AFOSR-TR- 84-0055

HFL-83-1

Quality Metrics Of Digitally Derived Imagery and Their Relation To Interpreter Performance: VIII.

Interim Report

AD A137726

30 June 1983

Harry L. Snyder, Ph.D.

Department of Industrial Engineering and Operations Research
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

Contract F49620-78-C-0055
Life Sciences Directorate
U.S. Air Force Office of Scientific Research
Bolling Air Force Base, D.C. 20332

HFL-83-1

Quality Metrics Of Digitally Derived Imagery and Their Relation To Interpreter Performance: VIII.

30 June 1983

Harry L. Snyder, Ph.D.

Department of Industrial Engineering and Operations Research
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

Contract F49620-78-C-0055
Life Sciences Directorate
U.S. Air Force Office of Scientific Research
Bolling Air Force Base, D.C. 20332

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A137 726	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) QUALITY METRICS OF DIGITALLY DERIVED IMAGERY AND THEIR RELATION TO INTERPRETER PERFORMANCE: VIII. ~~FINAL REPORT~~.		5. TYPE OF REPORT & PERIOD COVERED Interim Report 1 August 1978-30 June 1983
		6. PERFORMING ORG. REPORT NUMBER HFL-83-1
7. AUTHOR(s) Harry L. Snyder, Ph.D.		8. CONTRACT OR GRANT NUMBER(s) F49620-78-C-0055
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Industrial Engineering and Opers. Res. Virginia Polytechnic Institute and State University Blacksburg, Virginia 24061		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313/A4
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NL) Bolling AFB, D.C. 20332		12. REPORT DATE 30 June 1983
		13. NUMBER OF PAGES 94
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Digital imagery
Human factors
Image quality
Image interpretation
Image processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes a five-year research program on the subject of quality evaluation of digitally derived imagery. High-resolution aerial photography was used to create digitized images with unclassified content comparable to that employed in military photointerpretation operations. The digitized imagery was used in several hard-copy and soft-copy interpretation experiments to assess the effects of image blur and image noise on both perceived image quality and the ability to extract information from the images.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The soft-copy experiments included both nonprocessed and processed imagery. Finally, quality metrics of image quality were obtained for both hard-copy and soft-copy images and related directly to both information extraction performance and subjective quality scaling.

The results are consistent across the several experiments and indicate the following.

(1) The interpretation scenario developed in this program is consistent, useful, and operationally meaningful. It is recommended for use by researchers in this field to control irrelevant variables and to examine the effects of various processes and interpretation aids upon photointerpreter performance.

(2) There is a slight increase in information extraction performance with hard-copy imagery compared to soft-copy imagery, as used in this experiment. On the other hand, photointerpreters perceive the image quality of the soft-copy imagery to be slightly better than that of the hard-copy imagery. A novelty hypothesis tends to explain this result.

(3) Processing of the soft-copy imagery results in significant improvement of interpreter performance, overcoming the slight degradation of performance introduced by soft-copy display compared to hard-copy display. However, careful selection of the appropriate process is necessary, as some processes which are otherwise considered suitable can in fact degrade performance and subjective quality below that of a no-processing condition.

(4) Various quality metrics correlate extremely well, on a system basis, with photointerpreter performance and quality estimation. However, when such metrics are applied on an image-dependent basis, the prediction is not nearly as good, causing belief that meaningful weighting of various areas within a scene must be made in order to obtain image statistics on only interpretation-relevant areas.

A

A-1

PREFACE

This research was sponsored by Contract No. F49620-78-C-0055, from the Air Force Office of Scientific Research, under the monitorship of Dr. Alfred R. Fregly. Dr. Harry L. Snyder is the principal investigator.

The imagery used in this research was supplied by the Environmental Research Institute of Michigan (ERIM). Dr. James J. Burke and his staff at the University of Arizona scanned the original analog images to produce the digitized images which were used in this research. Mr. Gilbert Kuperman, AFAMRL, provided technical assistance throughout the effort. Colonel David Lehnertz, Commander, 460th Reconnaissance Technical Squadron, Langley Air Force Base, and his staff graciously and professionally supported data collection during the soft-copy phases, while the officers and staff of the 548th Reconnaissance Technical Group, Hickam Air Force Base, supported the hard-copy phases of the research. Mr. L. Hardy Mason and Mr. Charles D. Bernard of VPI & SU ably and creatively wrote the necessary software to make this research possible. Finally, a program of this magnitude is possible only with the dedication, hard work, and talents of many persons. For their excellent contributions, thanks are given to Mr. Robert J. Beaton, Mr. Thomas Bruegge, Mr. Kenneth R. Castle, Dr. Betty P. Chao, Dr. Michael E. Maddox, Mr. Ray Schmidt, Mr. David I. Shedivy, Dr. Robin N. Strickland, and Mr. James A. Turpin.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table **page**

I. INTRODUCTION

Recent technological developments have resulted in a wide variety of imaging systems and subsystems. The flexibility and technologies available to the system designer include various means for collecting, coding, transmitting, decoding, analog and digital processing, and analog and digital display. The applications of such systems and subsystems are myriad, ranging from static and dynamic military photointepretive functions, through commercial and closed-circuit television and facsimile systems, to diagnostic radiological instrumentation and earth resource applications. The scientific world is quite familiar with some of the techniques which can be used to "improve" the nature of any image, and the non-scientific world has been greatly impressed with examples of information enhancement through image processing.

In many cases, it is clear that image processing and display techniques can extract information in the original image that would otherwise be well below the threshold capacity of the human visual system, whereas in other cases it has been clear that processing techniques can often serve either to hide existing, and important, image detail or to "create" image detail that is perhaps not present in the original image or in the "real world." Heretofore, most of these areas of image system and subsystem development have plainly suffered from their inattention to human observer requirements. This is particularly true of the extensive effort in digital processing, especially that part devoted to the improvement ("enhancement",

"restoration") of images for the purpose of human information extraction. In nearly all of the work performed in laboratories around the country that are pursuing this type of research, the necessary evaluative efforts to determine the utility of processing and display techniques have not been conducted. Rather, reports and publications of this work have typically taken the form of "before and after" pairs of images, with which the reader is left to estimate the utility of such processing either by visual inspection of these published (second- or third-generation) photographs or by the subjective opinions offered in the text by the author.

Because the intent of such image processing techniques is to improve the information extraction capabilities of the human observer, it is clearly appropriate and mandatory that evaluative techniques include objective measurement of human information extraction from the images, in addition to subjective estimates of the overall quality or utility of the image. Unfortunately, the human factors experiments required to produce quantitative and objective assessment of image quality have rarely been conducted in image processing laboratories or in conjunction with image processing programs.

In view of the many millions of dollars being devoted to image collection, processing, and display systems for the military and civilian use of digitized images, it was quite clear that an assessment program was urgently needed to devise procedures, techniques, and metrics of digital image quality. That program required the establishment of a standardized set of procedures for obtaining human observer information extraction performance; relating that

performance, in quantitative manner, to the various collection, processing, and display techniques and algorithms; and devising a quantitative relationship for the effectiveness of various quality levels of digital imagery to collection, processing, and display techniques and parameters.

Only by such an integrated program of research can the system and subsystem designer have meaningful data for cost-benefit analyses of future system development, be such systems intended for military or for non-military purposes. Prior to the start of this present effort, the image collection, processing, and display technology had reached a point whereby such evaluative research was sorely needed. Fortunately, microphotometric, microdensitometric, and human performance measurement techniques had evolved during the past several years to permit relating human information extraction performance to the various physical characteristics of both electro-optical and photographic image displays. The present research program was therefore designed to extend these recently developed techniques into the arena of digital images, emphasizing derivation of metrics of image quality appropriate to digitized images, and providing quantitative performance data which permit the designer or system developer to plan his development effort as well as to specify optimum system components for particular image acquisition and display requirements.

OVERVIEW OF THE RESEARCH PLAN

The research plan is laid out schematically in Figure 1. Each

small, solid-lined box, with the exception of the uppermost, indicates a separate task that was conducted during the course of the five-year effort. The two large, broken-lined boxes delineate the specific display formats that were studied during the initial program: black-and-white hard-copy transparencies and electronic displays. The small, broken-lined box at the bottom illustrates important extensions of this research to be pursued, hopefully, in the future, namely interactive digital displays in both monochrome and full color.

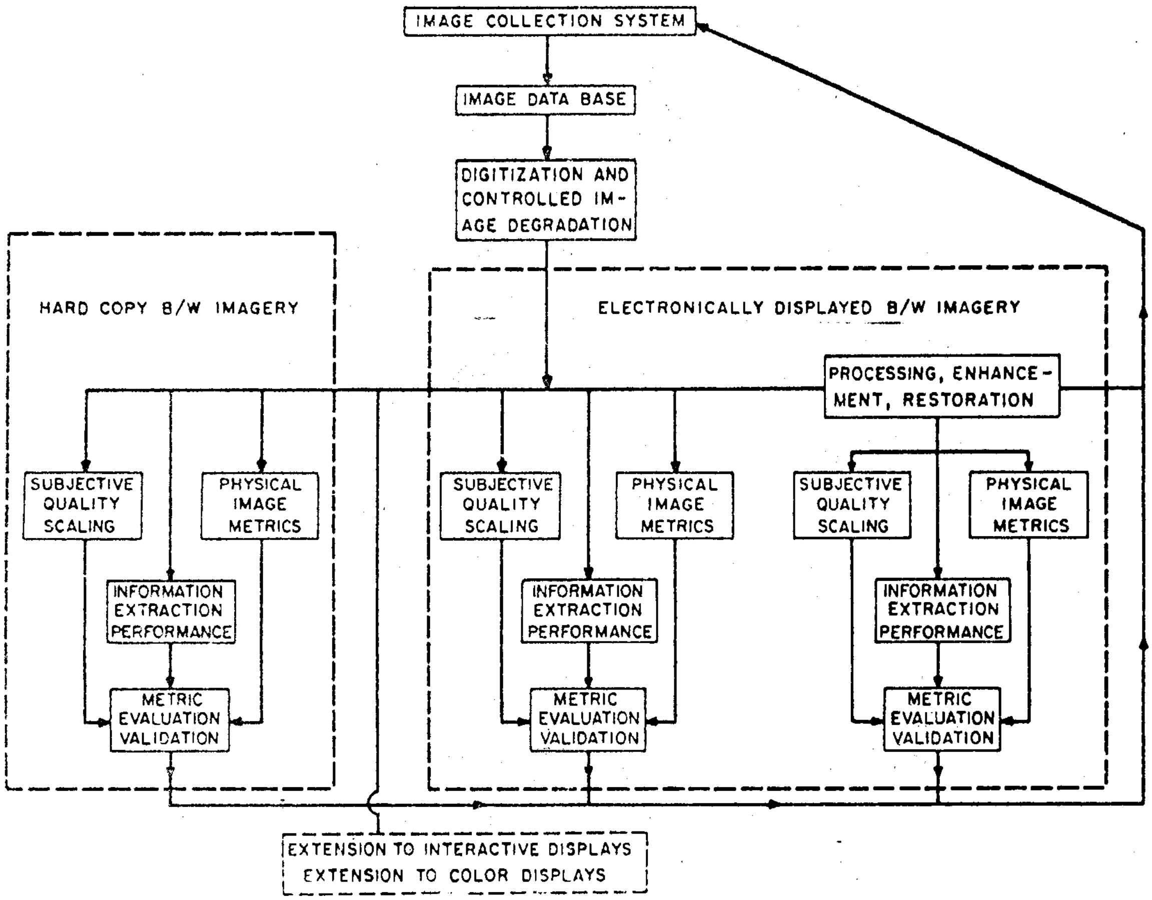

Figure 1. Schematic overview of the research program.

RESEARCH OBJECTIVES

The overall research objectives of this program were as follows:

1. Develop standardized procedures and techniques to evaluate hard-copy (film) and soft-copy (CRT) digital image quality.
2. Compare candidate physical metrics of image quality.
3. Compare hard-copy with soft-copy displays for image interpretation.
4. Evaluate candidate processing, enhancement, and restoration algorithms for improvement of image interpretation on soft-copy displays.

SPECIFIC RESEARCH TASKS

In keeping with the general goals listed above, the specific research tasks were as follows:

1. Develop an imagery database and image interpretation scenario from high quality aerial photography relevant to the image interpretation task.
2. Select and purchase display and interface hardware to present the image database on soft-copy displays.
3. Develop image manipulation software for soft-copy and hard-copy experiments.
4. Develop and standardize observer data collection procedures for hard-copy and soft-copy experiments.

5. Develop and standardize procedures for obtaining physical image quality metrics from hard-copy and soft-copy displays.
6. Digitize and degrade database imagery and record images on hard-copy and magnetic tapes for soft-copy experiments.
7. Obtain physical image metric data for hard-copy and soft-copy displays.
8. Conduct subjective quality scaling and information extraction studies on hard-copy images.
9. Conduct subjective scaling and information extraction studies on soft-copy displays.
10. Evaluate the utility of image quality metrics for both hard-copy and soft-copy imagery.
11. Conduct subjective scaling and information extraction studies on processed soft-copy imagery.
12. Compare image quality metrics for hard-copy and soft-copy images. Relate these results to concepts and models of human visual performance and to imaging system design variables.

This research program was begun in August, 1978 and was completed in June, 1983. This present report summarizes all the research results in the program; therefore, it serves as a type of detailed executive summary. The reader who is interested in the specifics of the results, methodologies, and database is urged to read the various technical reports and publications that deal with

individual research tasks. Those technical reports, archival publications, and conference papers are listed in Appendix A. Because this research was conducted in a university environment, the program had the added benefit of suppporting numerous students and staff. Those persons participating in the effort and the students who received graduate degrees by contributing to this program are listed in Appendix B.

II. METHODOLOGY

This research program required the controlled display of realistic images that were meaningful in an operational photointerpretation scenario. Therefore, early in the program, images were selected from the database at the Environmental Research Institute of Michigan (ERIM) which had the potential of meeting the experimental objectives. These images, in positive transparency form, were evaluated by senior photointerpreters (PIs) at the 460th RTS, Langley Air Force Base, for their realistic content and interpretative potential. Ten images were selected by these PIs for the subsequent experimental program. Each of the 10 images was subjected to the quantification and manipulation described below.

DATABASE PREPARATION

Operationally meaningful ranges of image blur and image noise were selected from a variety of images which had been processed to produce blur and noise over a much larger range. Based upon these early recommendations by the 460th RTS PIs, the final database images were designed to produce five levels of blur and five levels of noise, for a combined database of 25 image quality conditions for each of the 10 image scenes, or a total of 250 images. Each of the 10 image scenes was then processed, by the techniques described below, to generate 25 blur/noise conditions on a magnetic tape record.

The nominal (intended) values of both blur and noise were

initially described in the following terms. Blur was defined as the full-width half-maximum (FWHM) of the equivalent gaussian intensity distribution of the individual picture element. Blur levels were nominally set as 20, 40, 80, 160, and 320 micrometers on the original image size of approximately 7.6 x 7.6 cm. Since each image was composed of 4096 x 4096 picture elements (pixels), the center-to-center spacing of pixels was 20 micrometers, or the FWHM of the no-blur condition. Increasing amounts of blur were obtained by digitizing each image with the PDS microdensitometer at the University of Arizona Optical Sciences Center and digitally processing these 4096 x 4096 array images to produce the desired blur. Each image was digitally blurred by overlapping 9 x 9 $(512)^2$ fast Fourier transforms, each done "in place" on a large memory VAX 11/780 at the University of Arizona. The Fourier transforms were multiplied by four appropriate gaussian filter functions to create the four highest blur levels. The products were then inverse Fourier transformed to yield, after discarding overlap, the required images with the specified blur, these images then being written to magnetic tape for storage. The actual blur levels produced by this process, as recorded on magnetic tape, were very close to the nominal values--specifically, they were 22, 43, 81, 161, and 320 micrometers, again referenced to the original image format size.

The noise dimension was added after the completion of the blur process. Since each image had originally been "stretched" in contrast to yield a dynamic range of nearly 2000 levels (11 bits), noise was added to yield a known signal-to-noise ratio (SNR) for each image, where signal-to-noise is defined as the peak signal divided by the rms

noise, both expressed in digital values. The nominal values of SNR were 200, 100, 50, 25, and 12.5. These values were obtained quite accurately on the tape recorded images.

Hard-Copy Image Preparation

The creation of hard-copy transparency images from the magnetic tape data was performed by the Image Processing Institute, University of Southern California. Each 4096 x 4096 tape was played into a Dicomed Model D-47 to print an 86 x 86 mm image in negative transparency form, from which positive contact transparencies were produced by personnel at the Optical Sciences Center, University of Arizona. Due to both noise and resolution limitations of the Dicomed, this printing sequence resulted in different blur and SNR values than were originallly intended.

The resulting FWHM blur and noise values are defined hereafter in a somewhat different fashion to permit comparisons across the several hard-copy and soft-copy experiments that followed. Specifically, the blur dimension is defined in FWHM pixels, referenced to the original pixel size of 20 micrometers. In this fashion, the variable magnification selected by the interpreter for both hard-copy (microscope) viewing and soft-copy (electronic zooming) is disregarded. The noise dimension is defined as the square root of the area under the two-dimensional Wiener spectrum of the noise multiplied by the MTF of the display system, either hard-copy or soft-copy. That is, it is a measure of the RMS transmissivity of the inserted noise, as passed by the **entire** display system. Details of these blur and noise

calculations for both hard-copy and soft-copy images have been reported by Beaton (1983), while details of the preparation and quantification of the original database have been reported by Burke and Strickland (1982).

Soft-Copy Image Display

The data tapes prepared by the Optical Sciences Center, as described above, were used to generate soft-copy images on laboratory quality cathode-ray tube (CRT) monitors in the Human Factors Laboratory, Virginia Polytechnic Institute and State University. The tapes were mounted on a magnetic drive peripheral of a DEC PDP 11/55 minicomputer, transferred to 160 Mbyte Ampex disc drives, and accessed via custom software to present a 512 x 512 pixel image on a pair of Conrac QQA-17 monitors. All image control, processing, and conversion for display were performed by an International Imaging System (IIS) Model 70 Imaging System. Monitors were optimized for maximum modulation transfer function (MTF) response and were kept in calibration throughout the experiments. The right side monitor presented a global image, in which the scene was subsampled each eighth pixel both horizontally and vertically. The left side monitor was equipped with a trackball which permitted the PI to "roam and zoom" over the entire global image to select portions of the global at a magnification of 2:1, 4:1, and 8:1, the 8:1 being full resolution (no subsampling) of the original image. Custom software provided this capability.

Linearization of the monitors in terms of luminance output vs.

bit level input was achieved by look-up tables (LUTs) in the IIS system using input corrections from a calibrated and periodically checked radiometric measurement system. This radiometric system was also used to measure both the noise and the blur contributions of the IIS/monitor combination. A 25-micrometer slit microphotometer was scanned along a sample of four raster lines containing a constant gray level to measure the spatial noise contribution of the system, while the same slit aperture was used to obtain an edge scan of a single vertical line of an image on the display. The edge scan was differentiated and Fourier transformed to obtain the MTF of the IIS/monitor combination. This MTF was cascaded with the equivalent MTF of the tape images, to obtain the "system" MTF of the soft-copy displayed image for each blur level. This system MTF was then used to compute the equivalent FWHM of the soft-copy displayed image.

Summary of Blur and Noise Levels

Because the magnification of the soft-copy image as viewed by the PI was much greater than that of the hard-copy image, and also because both the hard-copy and soft-copy images could be viewed at various magnifications selected by the PI, it is more meaningful to think of the FWHM independently of the magnification level than in terms of the FWHM relative to the original pixel sampling size of 20 micrometers. Accordingly, Table 1 presents the FWHM values for both hard-copy and soft-copy in these units along with the nominal values.

In like fashion, it should be realized that the noise added to the image to form the magnetic tape image (and its nominal SNR) must

ultimately be passed by the MTF of the display medium, either hard-copy or soft-copy. In that process, the noise spectrum is attenuated by the bandpass or MTF of the display medium and altered, particularly in the high-frequency end. For this reason, the SNR as presented on tape is probably not the most meaningful or descriptive term for the noise result, but rather one should use a measure of noise power as displayed to the PI. One such appropriate noise power measure is the Weiner spectrum. Using the square root of the area under the cascaded two-dimensional Wiener spectrum as a measure of noise power in the full system (displayed) image, the noise levels for the hard-copy and soft-copy images, as compared to the nominal levels, are given in Table 2.

It is recognized that the values of Tables 1 and 2 are not the only units in which the noise and blur dimensions can be expressed; however, because of the system-oriented nature of this program, it is believed that they are the most useful to the systems designer. Furthermore, other measures of blur and noise can be derived and defined from these, as described by Beaton (1983).

TABLE 1. Nominal and Measured Values of Blur Used in this Research

Nominal Value (micrometers)	Hard-Copy Value (FWHM Pixels)	Soft-Copy Value (FWHM Pixels)
20	2.146	0.902
40	3.025	2.331
80	4.766	4.464
160	8.789	8.747
320	17.270	17.324

TABLE 2. Noise Levels Used in this Research

Signal-to-Noise	Displayed Wiener Spectrum, rms Transmissivity	
Nominal Value	Hard-Copy Value	Soft-Copy Value
200	0.00767	0.00582
100	0.00958	0.00936
50	0.01378	0.01578
25	0.02402	0.02997
12.5	0.04457	0.05786

HARD-COPY EXPERIMENTS

Fifteen military photointerpreters of the 548th Reconnaissance Technical Group, Hickam Air Force Base, Hawaii, served as subjects in these experiments. The same PIs performed in the first (information extraction) experiment and in the second (subjective quality scaling) experiment. One PI declined to participate in the second experiment, reducing the number of subjects in that experiment to 14.

Information Extraction

In the information extraction experiment, each PI received 10 images to evaluate by answering a series of specific questions regarding essential elements of information (EEIs) in the images. This task was designed to be quite similar to the daily interpretive tasks of the PIs in normal assignments. Five of the 15 PIs were randomly assigned to each of three blur levels (2.146, 4.766, and 17.270 pixels). Each PI viewed two scenes at each noise level (0.00767, 0.00958, 0.01378, 0.02402, and 0.04457 rms transmissivity), with the scenes presented at each noise level represented equally often across the five PIs in each blur condition. The order of each unique scene/noise combination was randomized for each PI.

Standard light tables and binocular zoom stereo optics were provided to the PIs. In addition, they were permitted to use any additional equipment of their choice. Standard photointerpretive reference volumes were provided to the subjects as aids in the task.

Pen and paper were used to record all answers to the EEI questions. No time limit was set on the task.

The EEIs were generated by a panel of senior PIs at the 460th Reconnaissance Technical Squadron, Langley Air Force Base, Virginia. Based upon the ground truth of the images, the answers to these EEIs were also determined and weights assigned for each possible partial answer. This **a priori** scoring scheme was used in this hard-copy experiment and in the subsequent soft-copy experiments. Scores were normalized by image, and a percent correct score for each image was determined for each PI. The percent correct scores provided the data for subsequent statistical analyses. Details of this methodology and the procedures followed are contained in the report by Snyder, Turpin, and Maddox (1982).

Subjective Quality Scaling

Fourteen of the 15 PIs who participated in the information extraction experiment also participated in this experiment, which followed immediately after the information extraction experiment for each PI. That is, in a typical week, a PI would participate in the information extraction experiment for two days and in the subjective quality scaling experiment for three days.

Each PI received, in individually randomized order, all 250 images (all combinations of noise, blur, and scene). The PI evaluated each image on the light table and assigned a quality rating based upon the NATO rating scale. On this scale, values range from zero (totally uninterpretable) to nine (which permits detailed analysis and

interpretation). To achieve greater resolution than would otherwise be possible with the 10-point scale, the PIs were instructed to expand the scale by using decimal values (e.g., 3.6, 7.4) to create a 100-point scale. The NATO Scale is shown in Appendix D as it was used in this study and in the subsequent soft-copy experiments.

SOFT-COPY EXPERIMENTS

The soft-copy experiments were conducted in a fashion very similar to the hard-copy experiments. The first soft-copy experiment evaluated information extraction performance while the second obtained subjective quality scaling data. The subjects for these experiments were PIs from the 460th RTS.

Information Extraction

Fifteen PIs were employed in this study, five assigned to each of three blur levels (0.902, 4.464, and 17.324 pixels). Each PI interpreted 10 images, one per scene, two of which were at each of the five noise levels (0.00582, 0.00936, 0.01578, 0.02997, and 0.05786 rms transmissivity). The same EEIs and scoring scheme were used as in the hard-copy study.

As described by Chao, Beaton, and Snyder (1983) the PI had a global image of the entire scene (at the apropriate blur, noise levels) on one 17-in. CRT and could command a subsection of that global image to the other 17-in. monitor. Cursor manipulation via a trackball and discrete button selection on the trackball unit permitted selection of

2:1, 4:1, or 8:1 magnification of the global image. All interpretation was performed from the "roamed and zoomed" image. Upon request from the PI, the experimenter would rotate the roamed and zoomed image 90 or 180 deg. Auxiliary information was the same as that used in the hard-copy study.

Subjective Quality Scaling

The same 15 PIs participated in the subjective quality scaling study, which was scheduled immediately following the information extraction experiment. Each PI used the 100-point NATO scale to assign a quality value to each of the 250 images (all combinations of scene, blur, and noise). The display of each global scene was provided as in the information extraction study, but minor modifications were necessary, in the interests of time, for the magnified images on the other monitor. Specifically, senior PIs from the 460th RTS selected between two and four subportions of each scene that were considered pertinent to the subjective scale value determination. For each of these subportions, the most suitable magnification was determined. These selectively magnified and selected subportions were then displayed on the second monitor under the control of the PI, who could sequentially select these several subportions until he or she was satisfied that a scale value could be reliably assigned. At that time, the scale value was reported verbally to the experimenter and the next image was displayed. As in the information extraction experiment, rotation of the image in 90 or 180 deg increments was performed by the experimenter at the request of

the PI.

PROCESSED SOFT-COPY EXPERIMENTS

Two processed soft-copy experiments were conducted to evaluate the effectiveness of digital image processing upon both information extraction and subjective quality. Ten different restoration/enhancement conditions were evaluated in the subjective scaling study, while five of these were used in the information extraction experiment. The 10 processing conditions, listed by the intended function of each, are:

Contrast Modification

1. linear stretch
2. adaptive contrast stretch + noise filter

Deblurring

3. unsharp masking + noise filter + linear stretch
4. Laplacian filter + noise filter + linear stretch

Noise Removal

5. noise filter
6. neighborhood averaging + linear stretch
7. adaptive noise filter + linear stretch

Deblurring and Noise Removal

8. Wiener filter + noise filter + linear stretch

Control Conditions

9. noise filter + linear stretch

10. no processing

Information Extraction

In this experiment, 10 PIs from the 460th RTS served as subjects to evaluate the effects of five enhancement/restoration conditions on images containing 10 combinations of blur and noise. The five processing conditions were noise filter + linear stretch, unsharp masking, adaptive contrast stretch, neighborhood averaging, and the Wiener filter (processes 9, 3, 2, 6, and 8 above, with linear stretch and noise filtering added as indicated in the above list). The experimental design was chosen on the basis of efficiency, namely two 5 X 5 Graeco-Latin squares in which each PI interpreted one image under a unique combination of scene, blur, noise, and process. The blur/noise combinations used in this study were the following, in which the first value is blur and the second is noise: 0.902/0.00582, 0.902/0.01578, 0.902/0.05786, 4.464/0.00582, 4.464/0.01578, 4.464/0.05786, 8.747/0.02997, 17.324/0.00582, 17.324/0.01578, and 17.324/0.05786. Details are given in the report by Chao (1983).

With the exception of the digital image processing, the procedures employed in this experiment were the same as those in the previous soft-copy information extraction study. Each PI had a global image on one CRT monitor and a selectable "roamed and zoomed" image on the other monitor. Image rotation was available by experimenter command. Answers to the EEI questions were manually recorded and scored in accordance with the previously established procedures and criteria. Processing time per subimage selected by

the PI took from 2 to 105 s, depending on the process, compared to 2 to 14 s for roamed and zoomed subimages under no-processing conditions.

Subjective Quality Scaling

Each of the 10 PIs who participated in the processed information extraction study also participated in the processed soft-copy subjective quality scaling experiment. Using the same NATO scale, each PI assigned a scale value to 450 images, composed of all combinations of scene (five were selected), all 10 processes, three blur levels (0.902, 4.464, and 17.324 pixels), and three noise levels (0.00582, 0.01578, and 0.05786 rms transmissivity). The order of presentation of the 450 images was randomized for each PI.

The procedure followed in this experiment was the same as in the previous soft-copy scaling experiment, except that the selected levels of magnification presented on the monitor were preselected on the basis of the blur level and the process so as to avoid aliasing of the image. Details of this limitation and the levels of magnification used are presented in the report by Chao (1983).

QUALITY METRIC EVALUATION

A major objective of this research program was to evaluate the validity of various candidate image quality metrics for digitally derived or presented imagery. This objective was met by accumulating a list of candidate metrics recommended in the literature by previous researchers and adding to that list several candidates derived in the

conduct of the current effort, measuring both hard-copy and soft-copy images to obtain values of those metrics for each image, and correlating the values of the metrics with both information extraction performance and subjective scaling values for both hard- and soft-copy modes of presentation. This major analysis effort extended over more than a year due to the measurement complexity and the size of the data arrays needed to calculate each of the metrics for each of the images. Details of the process and the resultant metric values are contained in the report by Beaton (1983).

Quality metrics were, in addition, calculated both as system metrics and as image-dependent metrics. System metrics are those designed to evaluate the metric for an imaging system as a whole and therefore are averaged over a number of images to predict the efficacy of the metric for predicting overall system performance. On the other hand, image-dependent metrics will have different values depending on the content of the image and are therefore designed to predict the PI's performance with a specific image based upon both system characteristics and statistics of the image itself.

Table 3 lists the 16 system metrics evaluated for both hard- and soft-copy experiments, while Table 4 lists the 20 image-dependent metrics which were evaluated. Details of the derivation, rationale, and calculational formulae for each of these metrics are given by Beaton (1983).

TABLE 3. System Image Quality Metrics Evaluated in Both Hard-Copy and Soft-Copy Experiments

Metric Abbreviation	Metric Name
EP	Equivalent Passband
PEP	Perceptual Equivalent Passband
IR	Intensity Ratio
PIR	Perceptual Intensity Ratio
SSF	Squared Spatial Frequency
PSSF	Perceptual Squared Spatial Frequency
EW	Equivalent Width
PEW	Perceptual Equivalent Width
MTFA	Modulation Transfer Function Area
GSFP	Gray Shade Frequency Product
ICS	Integrated Contrast Sensitivity
VC	Visual Capacity
Q3	Hufnagel's Q3 Metric
SN	Signal-to-Noise Ratio
PMQ	Perceived Modulation Quotient
IC	Information Content

TABLE 4. Image-Dependent Quality Metrics Evaluated in Both Hard-Copy and Soft-Copy Experiments

Metric Abbreviation	Metric Name
EP	Equivalent Passband
PEP	Perceptual Equivalent Passband
IR	Intensity Ratio
PIR	Perceived Intensity Ratio
SSF	Squared Spatial Frequency
PSSF	Perceived Squared Spatial Frequency
EW	Equivalent Width
PEW	Perceived Equivalent Width
MTFA	Modulation Transfer Function Area
GSFP	Gray Shade Frequency Product
ICS	Integrated Contrast Sensitivity
VC	Visual Capacity
Q3	Hufnagel's Q3 Metric
PMR	Perceived Modulation Ratio
PMQ	Perceived Modulation Quotient
IC	Information Content
MSE	Mean Square Error
PMSE	Perceptual Mean Square Error
IF	Information Fidelity
SC	Structural Content
CQ	Correlational Quality

IV. RESULTS

HARD-COPY SUBJECTIVE QUALITY SCALING

Details of the statistical analyses of the hard-copy subjective quality scaling experiment were reported by Snyder, Shedivy, and Maddox (1982) and are therefore not repeated here. As expected, increases in blur and noise reduced significantly the mean NATO scale value. As illustrated in Figure 2, the blur levels resulted in a variation in mean NATO scale value from nearly 6 (blur = 2.146 pixels) to approximately 3.3 (blur = 17.270 pixels). The effect is essentially linear.

The effect of image noise is illustrated in Figure 3, which indicates a reduction in mean NATO scale value from 5.3 (rms transmissivity of 0.00767) to 4.3 (rms transmissivity of 0.04457). The noise effect is also quite linear, but with a smaller range of variation in NATO scale values than was obtained over the blur levels.

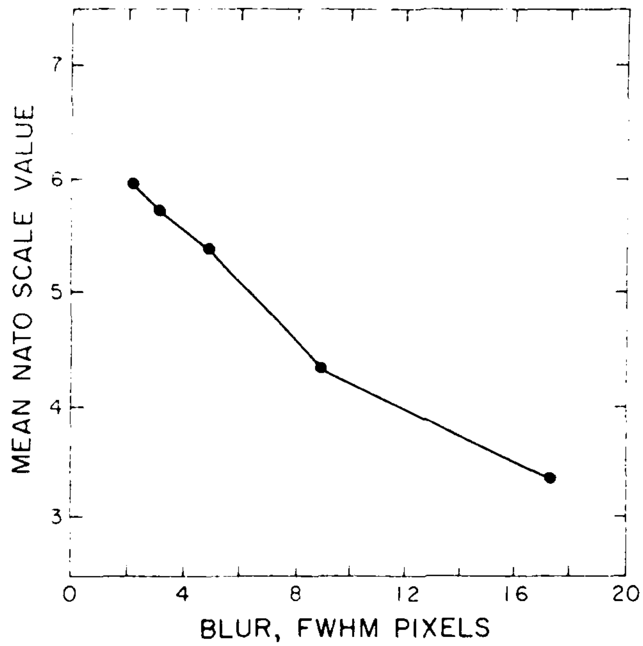

Figure 2. Effect of blur on mean NATO scale value for hard-copy imagery.

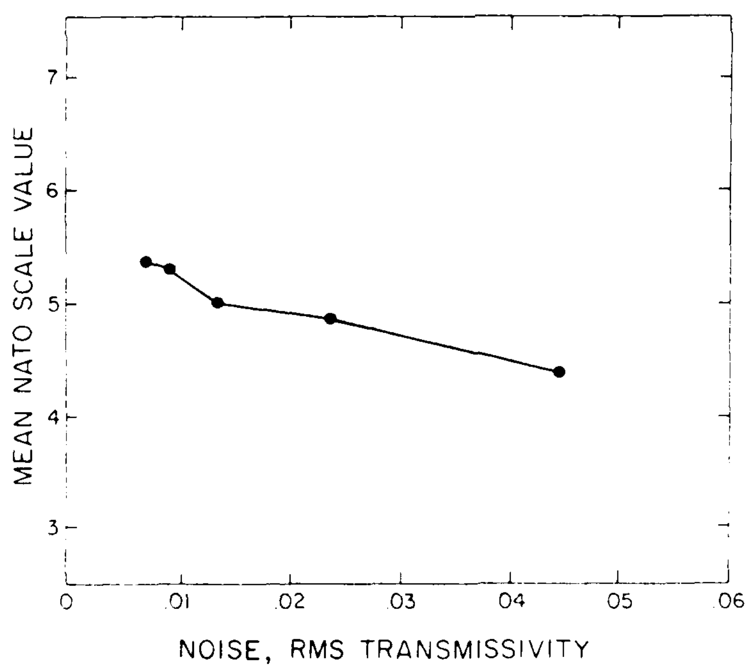

Figure 3. Effect of noise on mean NATO scale value for hard-copy imagery.

The noise X blur interaction, while statistically significant ($p < .001$), shows only a small contribution of blur to the noise effect (Figure 4). At the higher blur levels (17.270 and 8.789 pixels), the slope of the noise curves is less than at the lower blur levels. That is, with large amounts of blur, the noise effect is somewhat less pronounced. Conversely, with less blur in the images, the effect of noise on perceived quality is greater. In general, larger amounts of image degradation, caused by either blur or noise, tend to mask out the degradation effects of the other variable.

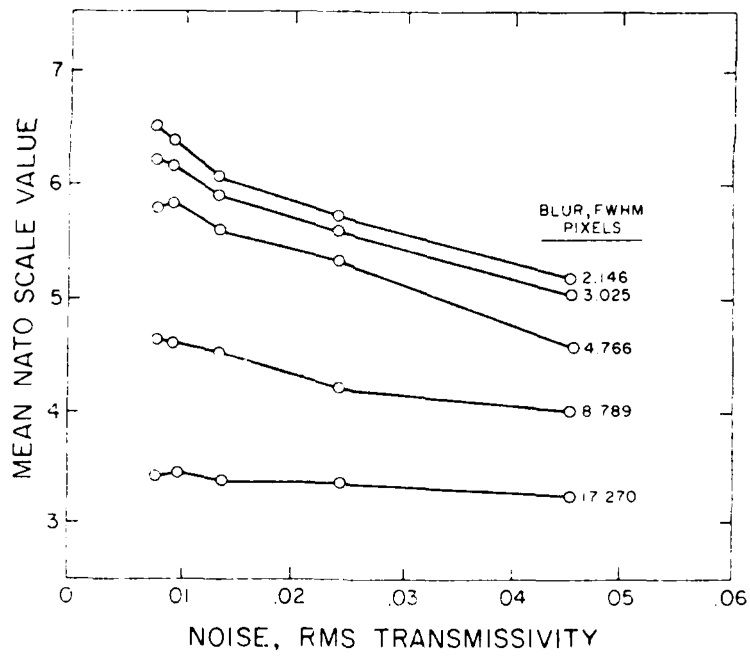

Figure 4. Effect of the blur X noise interaction on mean NATO scale value for hard-copy imagery.

HARD-COPY INFORMATION EXTRACTION

As illustrated in Figure 5, information extraction performance tends to follow the same pattern as do the mean NATO scale values. Increases in blur cause essentially linear decreases in the percent correct EEI score, although this effect is not statistically significant.

Increases in image noise likewise cause decreases in percent correct EEIs, as illustrated in Figure 6, except for a slight inversion between the lowest (0.00767 rms) and next lowest (0.00958 rms) noise levels; however, this inversion is not statistically significant. Disregarding the inversion of these two values, the effect of noise on EEI performance is quite linear.

Finally, the blur X noise interaction, illustrated in Figure 7, shows the same general trend that was observed for the NATO scale values. The effect of noise on percent correct EEIs appears to be slightly greater at low blur levels than at higher blur levels, although this interaction is not statistically significant (largely due to the nature of the experimental design, as discussed by Snyder, Turpin, and Maddox (1982).

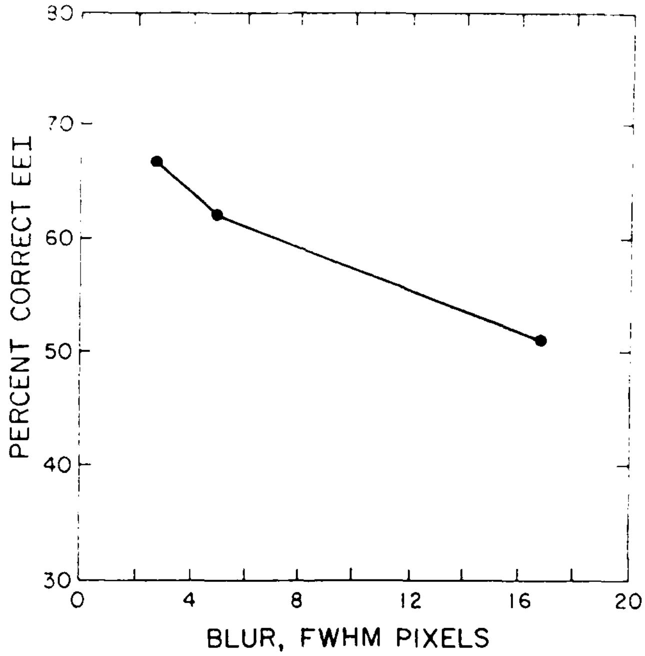

Figure 5. Effect of blur on information extraction performance for hard-copy imagery.

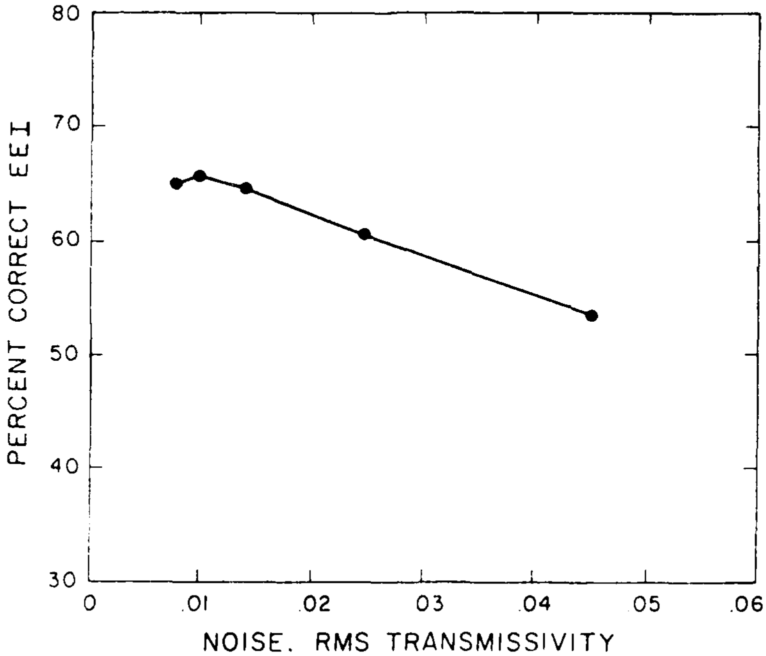

Figure 6. Effect of noise on information extraction performance for hard-copy imagery.

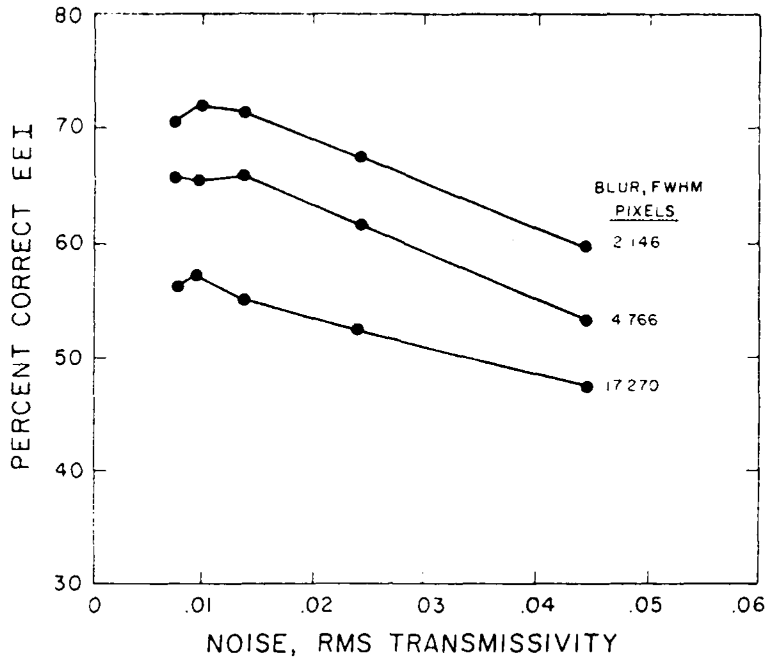

Figure 7. Effect of the blur X noise interaction on information extraction performance for hard-copy imagery.

SOFT-COPY SUBJECTIVE QUALITY SCALING

The effect of blur on the mean NATO scale value for soft-copy imagery is very similar to that shown above for hard-copy imagery. Increases in blur cause nearly linear decreases in NATO scale values, as illustrated in Figure 8.

In a similar fashion, increases in the noise content of soft-copy imagery result in consistent, nearly linear decreases in NATO means, as shown in Figure 9. Figure 10 indicates the nature of the

significant blur X noise interaction for soft-copy subjective quality. As is the case with hard-copy imagery, higher blur levels cause a reduction in the effect of noise on perceived quality. At the blur levels of 17.324 and 8.747 pixels, the curves are much flatter than for the lower blur levels. Again, increases in either noise or blur tend to mask the influence of the other variable on perceived quality.

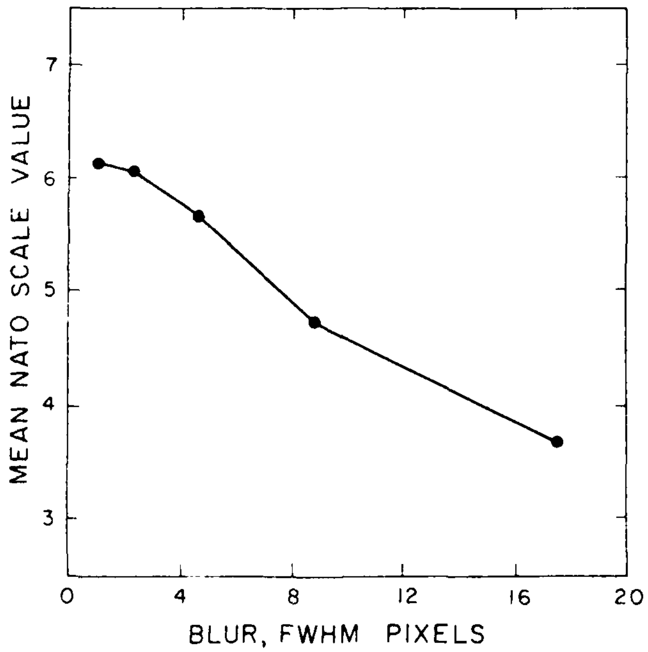

Figure 8. Effect of blur on mean NATO scale value for soft-copy imagery.

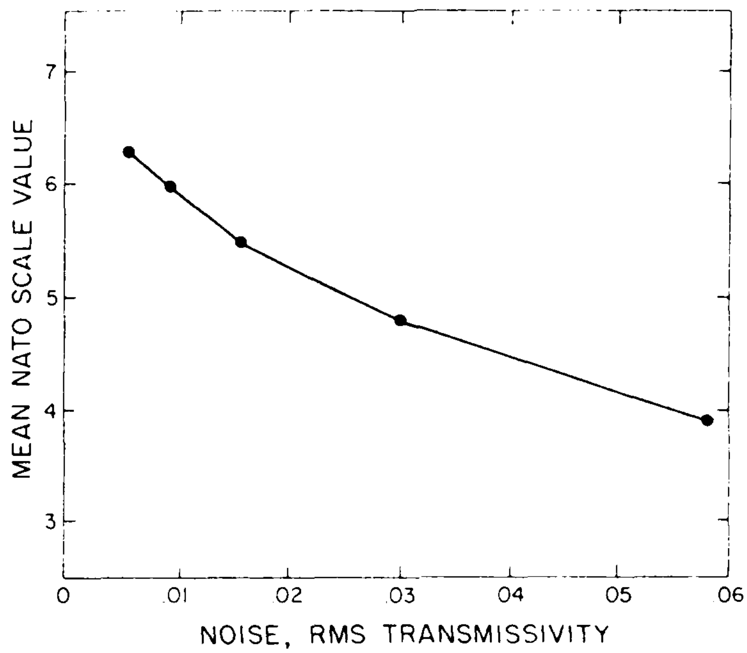

Figure 9. Effect of noise on mean NATO scale value for soft-copy imagery.

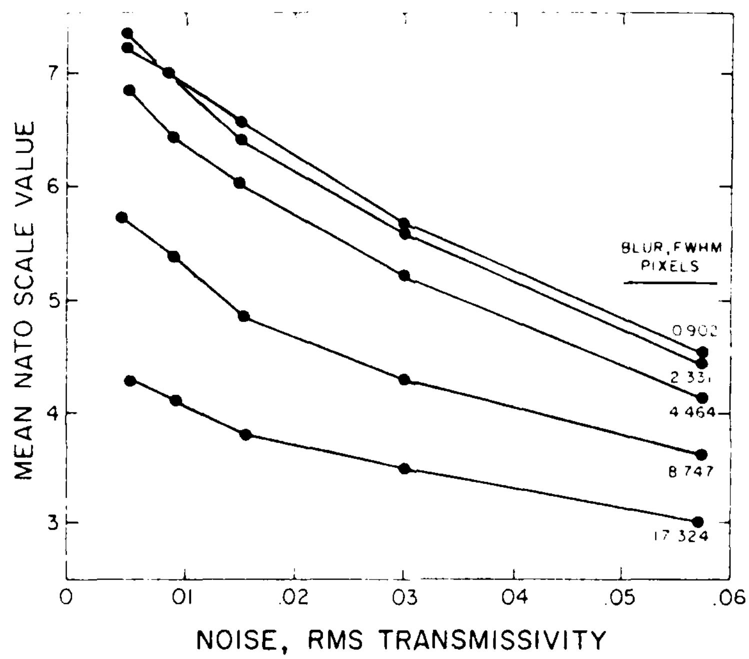

Figure 10. Effect of the blur X noise interac'ion on mean NATO scale value for soft-copy imagery.

SOFT-COPY INFORMATION EXTRACTION

The soft-copy information extraction results are quite similar to those of the hard-copy study. Figure 11 illustrates that the effect of blur is to cause a consistent reduction in percent correct EEIs, while Figure 12 indicates that increases in noise generally result in a reduction in EEI satisfaction. The small inversion in performance between the two lowest noise levels in Figure 12 is not statistically significant.

Lastly, Figure 13 shows the blur X noise interaction effect on information extraction for the soft-copy experiment. Again, there is a suggestion that the effect of noise on information extraction is somewhat less at the greatest blur level (17.324 pixels), although this interaction is not statistically significant ($p > .05$).

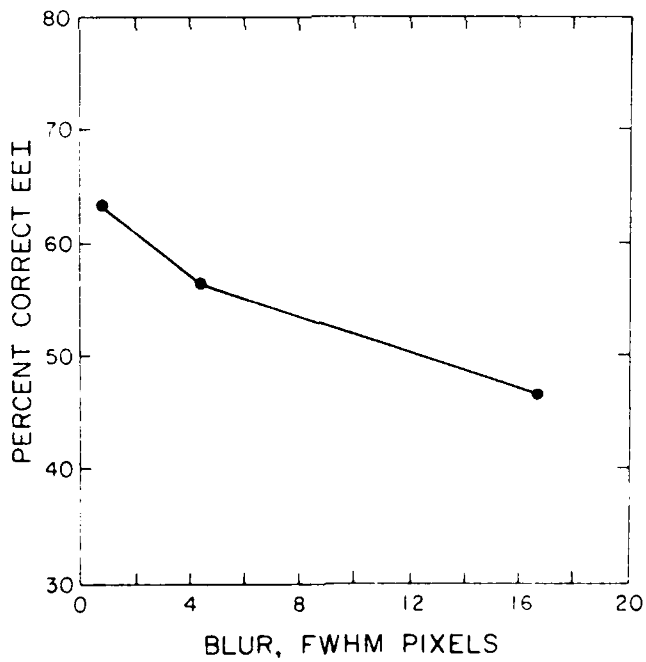

Figure 11. Effect of blur on information extraction performance for soft-copy imagery.

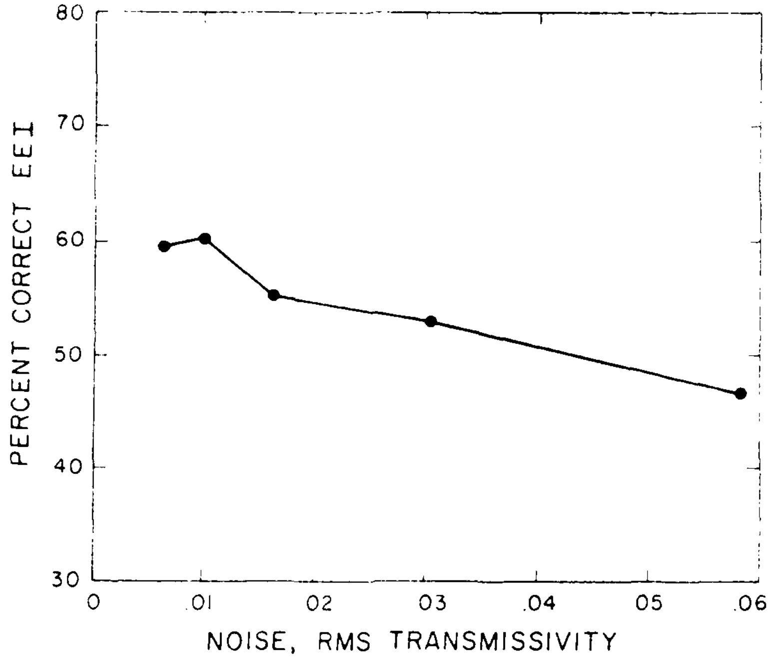

Figure 12. Effect of noise on information extraction performance for soft-copy imagery.

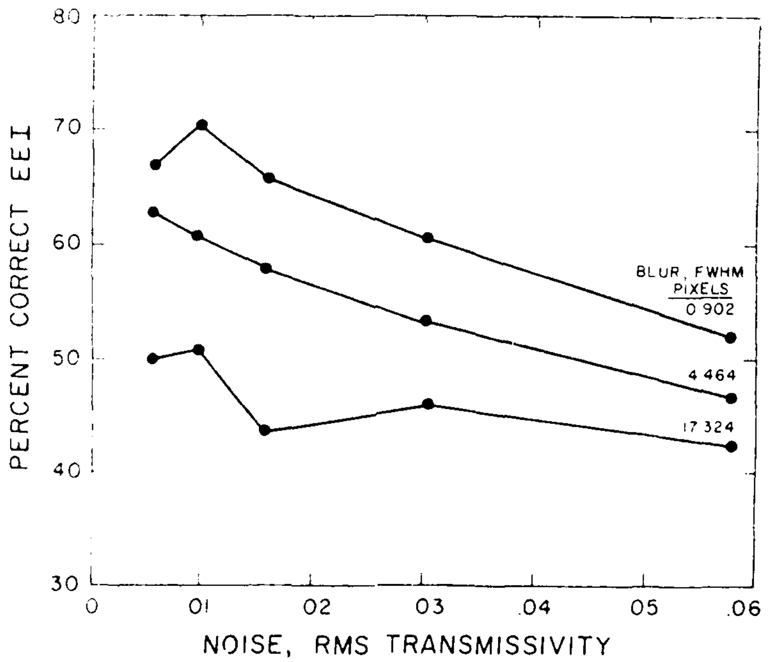

Figure 13. Effect of the blur X noise interaction on information extraction performance for soft-copy imagery.

COMPARISON OF HARD-COPY AND SOFT-COPY RESULTS

Subjective Image Quality

The effect of blur on perceived image quality is very similar for the hard-copy and the soft-copy experiments. As shown in Figure 14, however, the soft-copy imagery was perceived to be consistently better, for the same blur content, than was the hard-copy imagery. On the average, this difference is about 0.3 scale value.

In a somewhat different fashion, the effect of noise on perceived image quality is different for the hard- vs. soft-copy experiments. At the lower noise levels, the soft-copy was perceived to be of better quality than was the hard-copy imagery (Figure 15). However, as noise increases, the two functions converge, such that the perceived quality at an rms noise level of 0.045 is the same for both hard-copy and soft-copy imagery. Whether this trend would continue and hard-copy imagery would be perceived of higher quality than soft-copy imagery at greater noise levels cannot be determined reliably from these data.

The blur X noise interactions for the hard-copy and soft-copy experiments are shown in Figure 16. The general trends of the interactions are, of course, the same, with the effect of noise decreasing with increasing image blur for both types of imagery. In addition, at noise levels below 0.045, the NATO scale values for hard-copy are consistently below those for soft-copy. However, when the

hard-copy noise reaches the 0.045 rms transmissivity level, all five NATO mean values exceed the straight-line interpolated soft-copy values. Thus, the trend seen in the noise main effect is repeated for all blur levels.

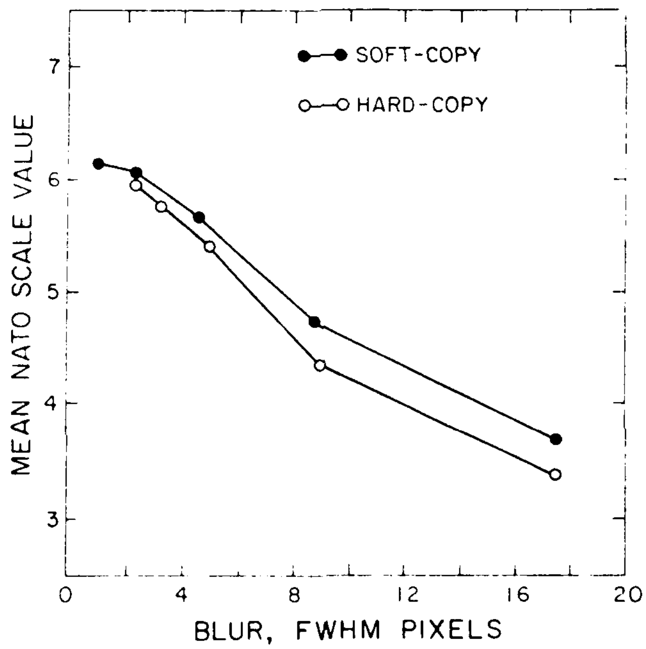

Figure 14. Effect of blur on mean NATO scale value for both hard-copy and soft-copy imagery.

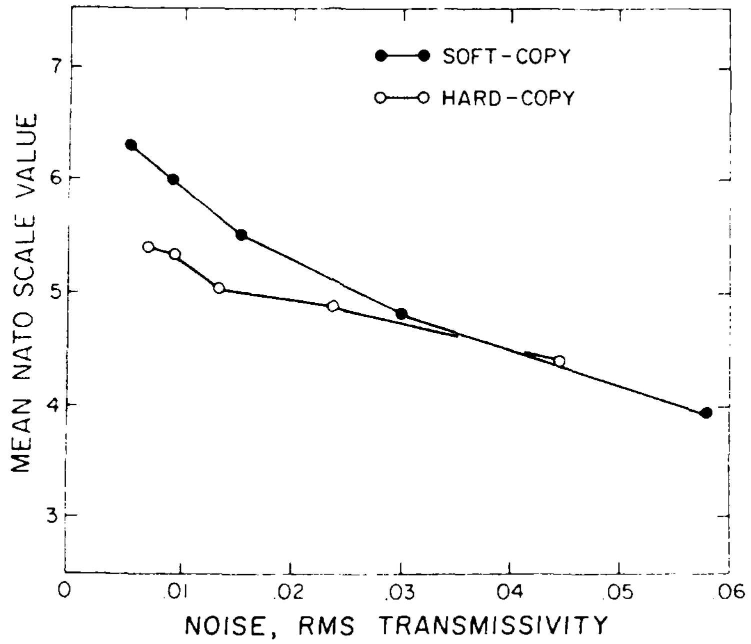

Figure 15. Effect of noise on mean NATO scale value for both hard-copy and soft-copy imagery.

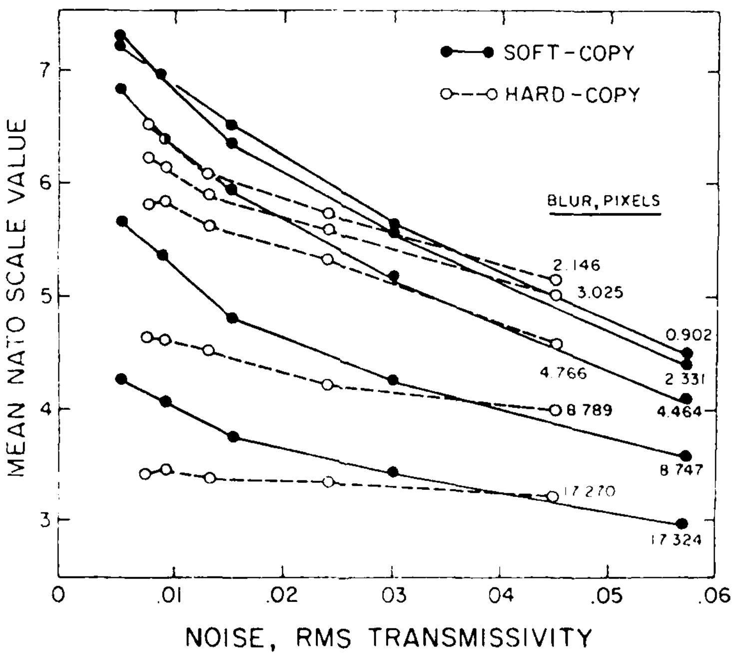

Figure 16. Effect of the blur X noise interaction on mean NATO scale value for both hard-copy and soft-copy imagery.

Information Extraction Performance

Whereas image blur produced greater perceived image quality for soft-copy than for hard-copy imagery (Figure 14), actual information extraction performance was **poorer** for soft-copy than for hard-copy imagery at all blur levels (Figure 17) and all noise levels (Figure 18). That is, while the perceived quality was consistently better with soft-copy, the actual EEI answers were inferior with the soft-copy presentation.

Furthermore, this result is consistent for all combinations of blur and noise, as illustrated in Figure 19. In each of the 15 blur, noise combinations, the percent correct EEI mean is greater for the hard-copy than for the soft-copy, a result which is highly statistically significant ($p < .0001$).

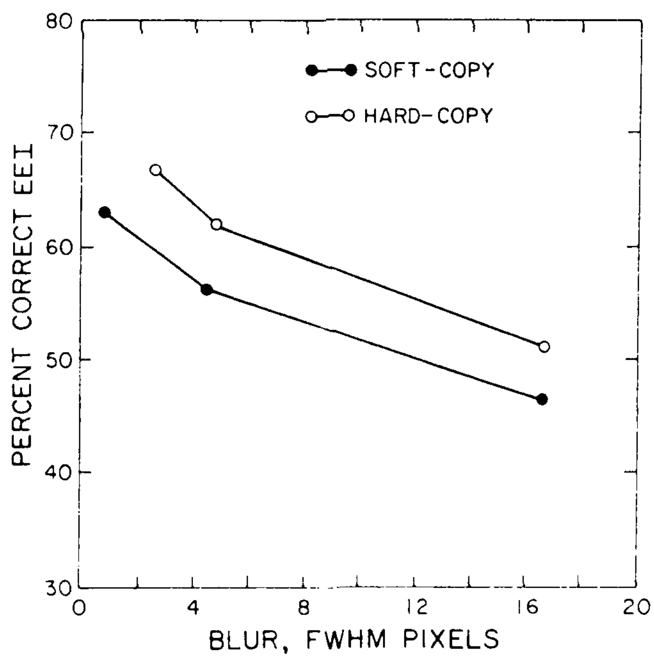

Figure 17. Effect of blur on information extraction performance for both hard-copy and soft-copy imagery.

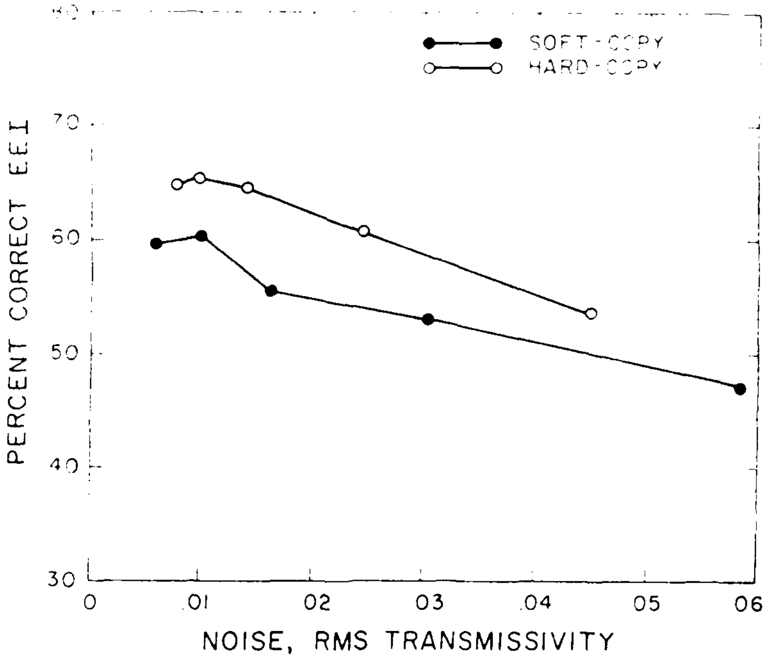

Figure 18. Effect of noise on information extraction performance for both hard-copy and soft-copy imagery.

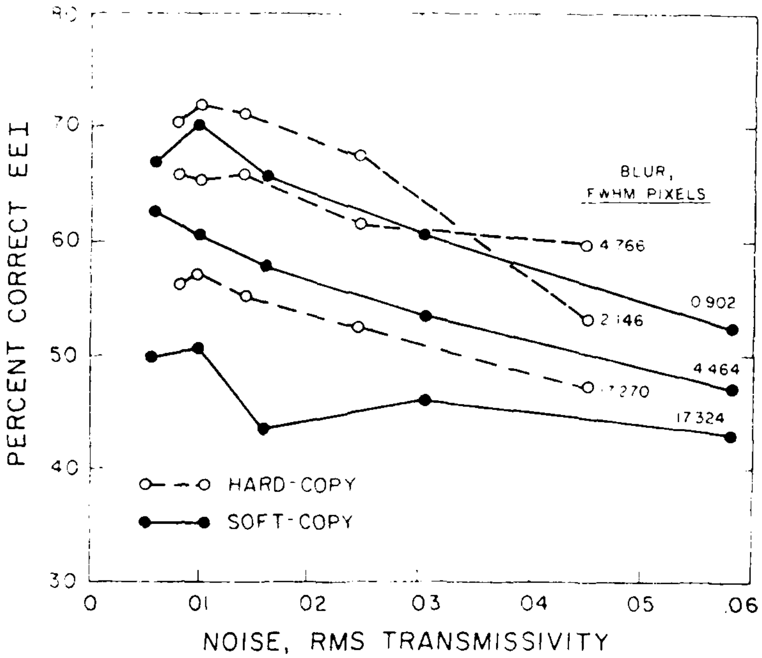

Figure 19. Effect of the blur X noise interaction on information extraction performance for both hard-copy and soft-copy imagery.

PROCESSED SOFT-COPY SUBJECTIVE QUALITY SCALING

Details of the analyses of the data from the processed soft-copy experiments have been reported by Chao (1983) and are only summarized here. Because the processed soft-copy subjective scaling experiment contained a control (no-processing) condition, the most meaningful way to evaluate the scaling results is to compare the mean NATO scale values for the process under consideration with the no-processing condition. In Figures 20 through 28, mean values are plotted for combinations of blur and noise for both the process and the no-processing control condition. Interpretations of these results are given below.

Contrast Modification

Two processes were selected to modify the image's contrast without altering the blur or noise content of the images. Figure 20 illustrates the effect of the linear stretch process on mean NATO scale values. In general, this process had little or no effect, with the only appreciable differences occurring at the high noise level where the process decreased the mean scale value for the lowest and middle blur conditions. Since the image database was "stretched" somewhat in its original preparation (Burke and Strickland, 1982), it is not unexpected that this additional stretching process had little influence.

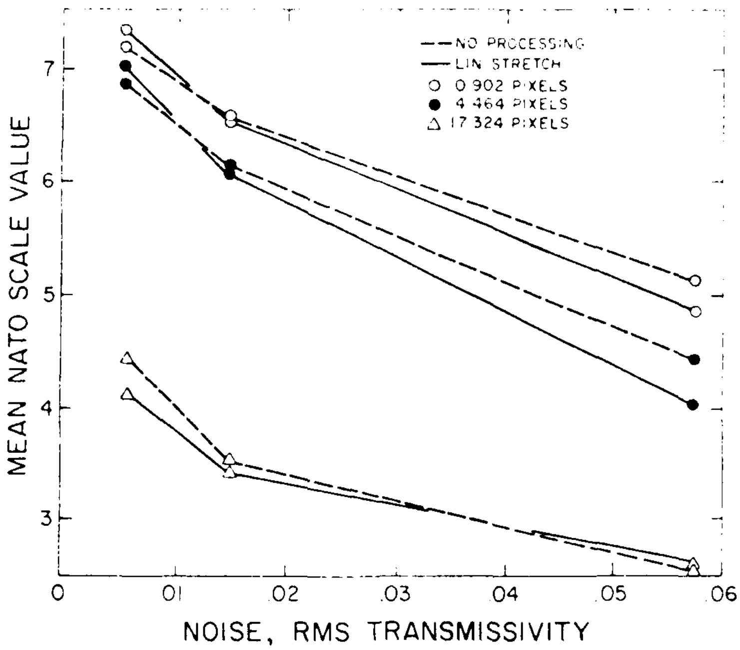

Figure 20. The effect of the linear stretch process on mean NATO scale values.

The second process, illustrated in Figure 21, was a combination of an adaptive contrast stretch plus a noise filter. Under low noise conditions (0.00582 rms transmissivity), this process reduced the perceived image quality, while under the high noise conditions (0.05786 rms transmissivity) the process improved perceived image quality. Thus, the adaptive contrast stretch component of this process probably added little to perceived quality, but the coupled noise filter appreciably improved quality when noise was present in any significant quantity. However, when noise was essentially absent, the process consistently caused a perception of reduced image quality. At the intermediate noise level (0.01578 rms transmissivity), there was no appreciable effect of this filter. Of

note is the fact that the average increase in NATO scale value at the highest noise level, due to the filter, is 0.7 scale unit.

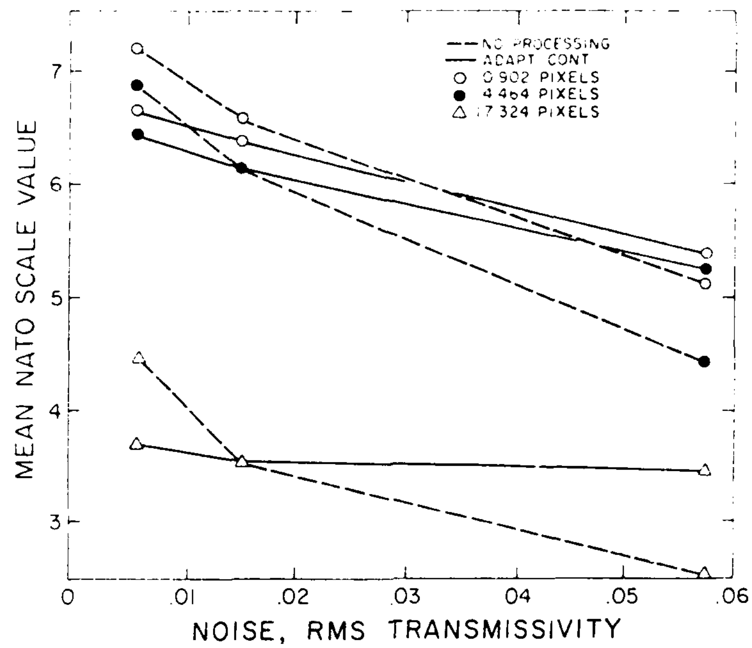

Figure 21. The effect of the adaptive contrast stretch + noise filter + linear stretch process on mean NATO scale values.

Deblurring Processes

Two deblurring processes were investigated, an unsharp mask and the Laplacian filter, each accompanied by the noise filter and a linear stretch. The noise filter was added because the application of the deblurring process alone would likely result in increased noise and this added noise would, under typical operational circumstances, have to be removed.

Figure 22 illustrates the influence of the unsharp masking + noise filter + linear stretch process on mean NATO scale values. For high blur imagery (17.324 pixels), the process improved perceived quality at the middle and highest noise levels. For the low and medium blur conditions, the process resulted in lower subjective quality. At the lowest noise level, regardless of the amount of blur in the image, there was essentially no effect of the process on subjective image quality. Thus, it appears that this particular process is useful only if there is significant noise **and** appreciable blur in the image, but that it is more harmful than useful if the image contains only blur or noise.

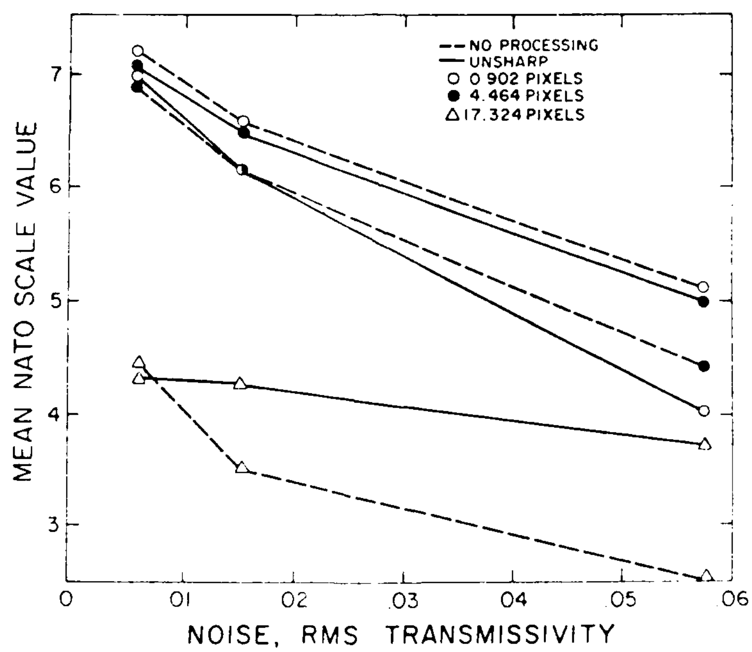

Figure 22. The effect of the unsharp masking + noise filter + linear stretch process on mean NATO scale values.

The Laplacian filter (coupled with a noise filter and linear stretch) produced very unusual results, as illustrated in Figure 23. Under high (17.324 pixels) or medium (4.464 pixels) blur, the process proved advantageous, particularly with high blur and high noise content. However, the low blur (0.902 pixel) images were viewed as having degraded image quality with the application of this filter. Under the highest blur condition, the average increase with this process was 0.6 scale unit, but the degradation caused by the process with the lowest blur images exceeded 1.4 scale units! It seems clear that this process must be applied with caution, as the disadvantages may often outweigh the advantages under some conditions.

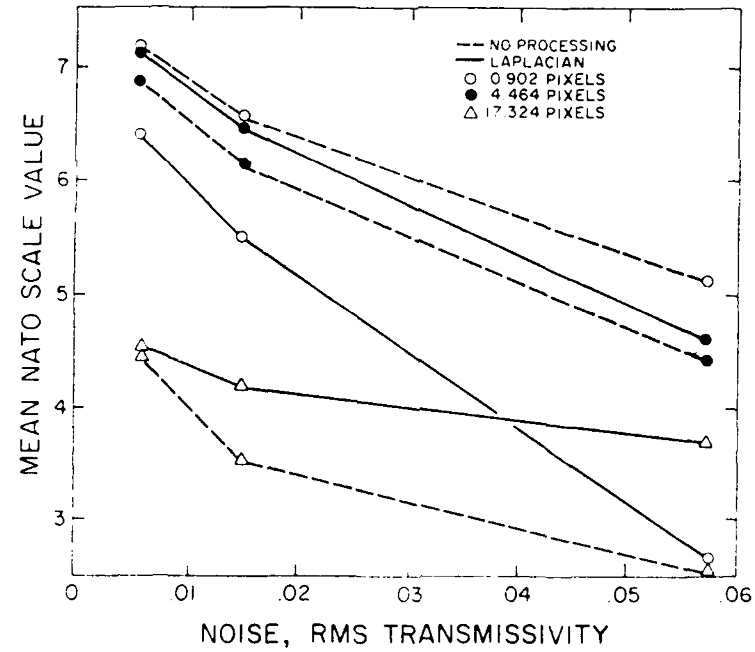

Figure 23. The effect of the Laplacian filter + noise filter + linear stretch process on mean NATO scale values.

The noise filter was effective at the highest noise level, regardless of the blur content of the images, resulting in an average scale increase of 0.8 unit, as illustrated in Figure 24. At the lowest noise level, there was essentially no effect, while there was consistent improvement at the intermediate noise level, averaging about 0.5 unit on the NATO scale. Only with the high blur, low noise images was there an average reduction in scale value, and this reduction was very small (0.3 unit). Thus, this process appears to be quite "safe" in improving the subjective quality of noisy images, with or without blur.

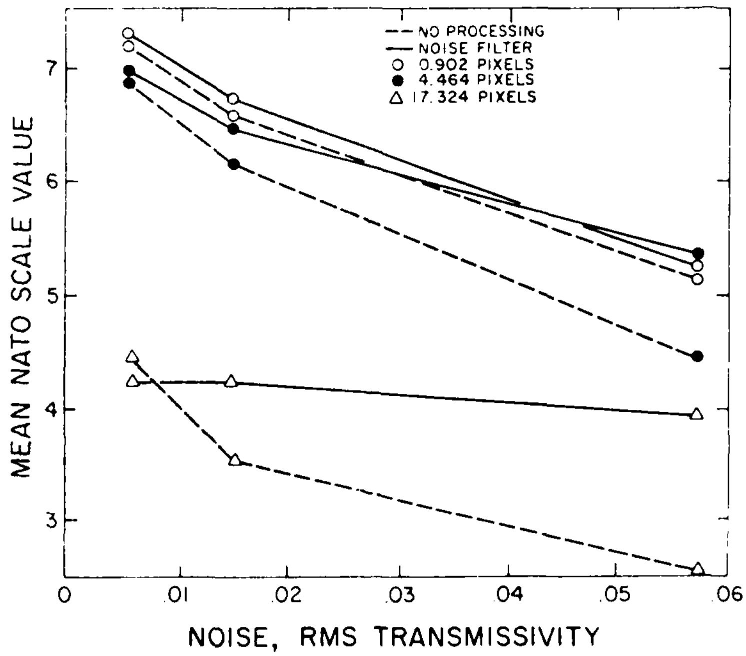

Figure 24. The effect of the noise filter process on mean NATO scale values.

The second noise removal process that was investigated was the neighborhood averaging + linear stretch combination. As indicated in Figure 25, this process improved subjective image quality consistently at both medium and high noise levels. However, at the lowest noise level, there was a reduction in image quality for the medium and high blur images. The lowest blur, lowest noise images were essentially unaffected by the process. The general results therefore indicate that the process is strongly recommended for images having noise in excess of 0.015 rms transmissivity, with the advantage averaging 0.8 scale unit at noise levels on the order of 0.058 rms transmissivity.

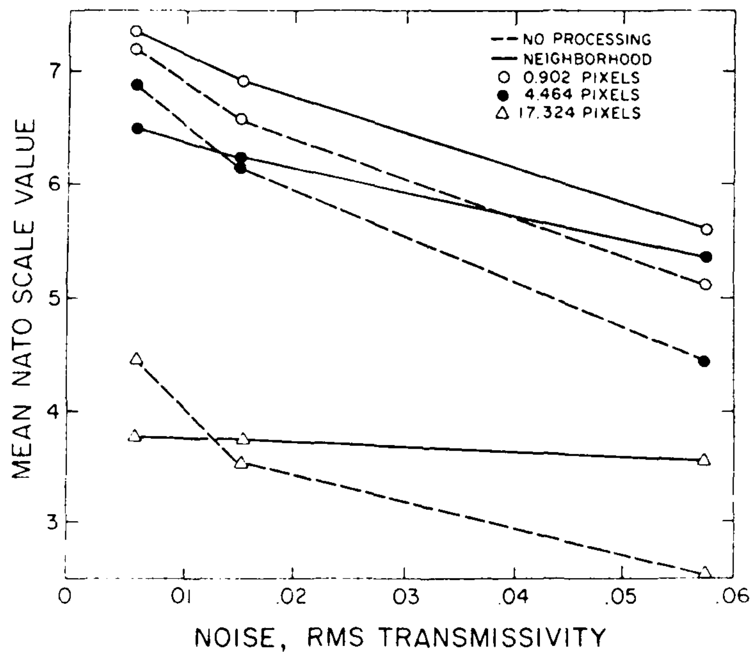

Figure 25. Tı effect of the neighborhood averaging + linear stretch process on mean NATO scale values.

The last noise removal process to be evaluated is the adaptive noise filter + linear stretch combination, the results for which are shown in Figure 26. With this process, as with the immediately preceeding process, the major benefits occur with noise levels in excess of about 0.03 rms transmissivity. At the highest noise level, the average improvement is about 0.75 NATO unit, while the improvement at the intermediate noise level occurs only with images having medium or substantial blur. At the lowest noise level (0.00582 rms transmissivity), there was a consistent reduction in subjective quality, regardless of the image blur. Thus, this process, while helpful in improving quality with noisy images, should not be applied to very-low noise images.

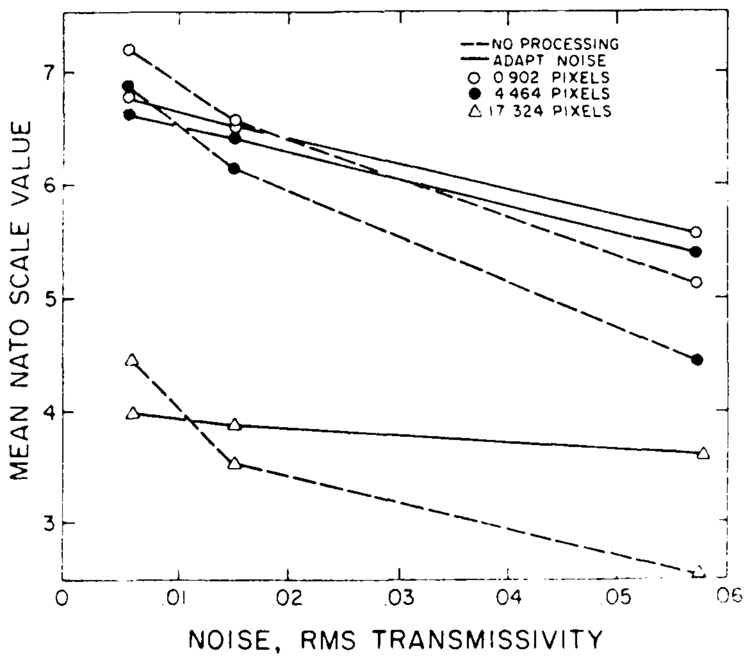

Figure 26. The effect of the adaptive noise filter + linear stretch process on mean NATO scale values.

The only filter in this experiment designed to both remove noise and to reduce blur in an image is the Wiener filter. Figure 27 illustrates the effect of this filter with a noise filter and linear stretch on the mean NATO scale values. At the lowest noise level, the process had little effect on subjective image quality. However, it increased the NATO mean value an average of 0.4 at the intermediate noise level and 0.8 at the highest noise level. Thus, the process was generally beneficial with little or no adverse effect under low noise or low blur conditions. Of course, this process requires some knowledge of the statistics of the particular image being processed, and therefore it is more complex to apply. Whether the cost/benefit tradeoff is positive can be argued.

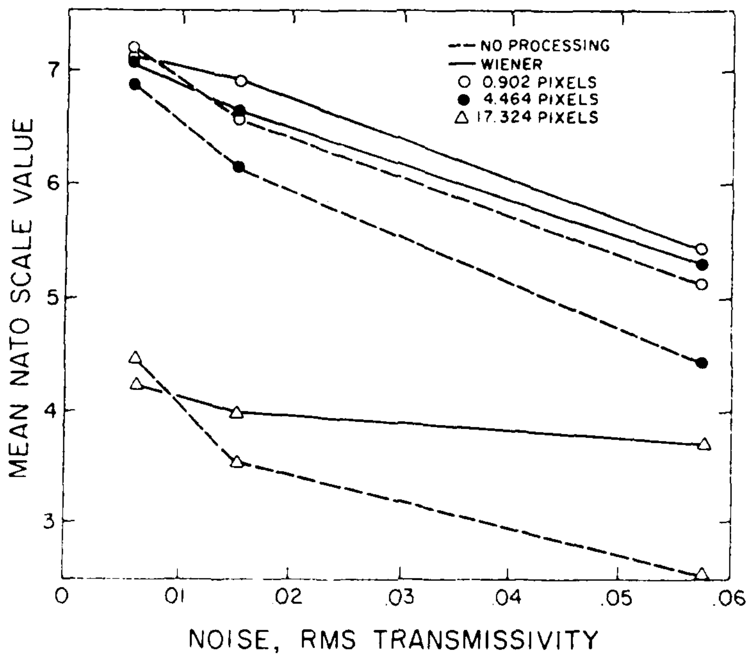

Figure 27. The effect of the Wiener filter + noise filter + linear stretch process on mean NATO scale values.

Control Condition

As noted previously, many of these processes must have inclusion of a linear stretch and/or noise filter to compensate for contrast attenuation and noise insertion in the processing. As a result, it was considered desirable to determine the influence of the noise filter and linear stretch components, per se, to the improvement achieved with any of the processes. The linear stretch + noise filter control condition is compared with the no-processing condition in Figure 28. Interestingly enough, the results are mixed. At the highest noise level, this "control" process improves the mean NATO value more than one unit for high-blur images and about 0.8 unit for intermediate blur images. However, for low blur images, the combination yields lower NATO scale values. At intermediate level noise, the combination helps the high blur image but harms the intermediate blur image. At low noise levels, there is little or no effect regardless of the blur level of the image. It seems clear, nonetheless, that some of the influence of the processes described above may well be due to the influence of the added components of linear stretch and noise filtering which followed the fundamental process application.

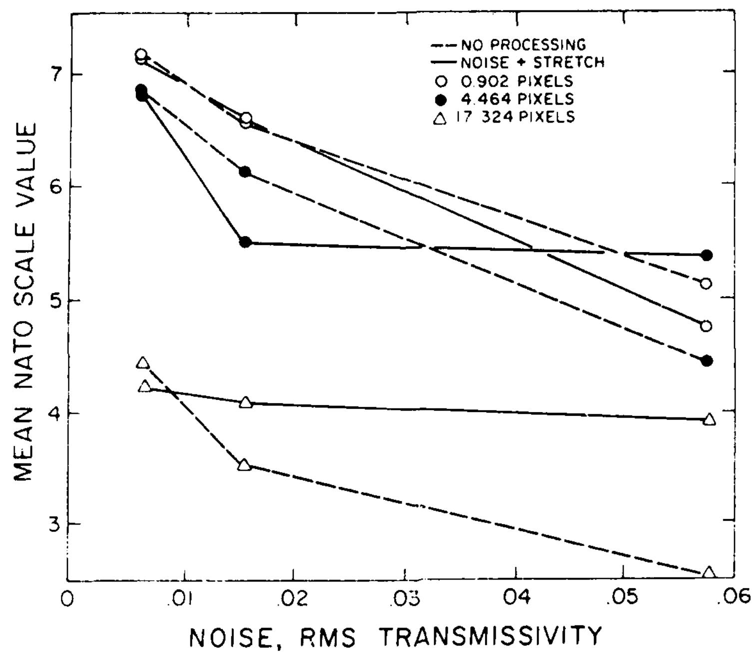

Figure 28. The effect of linear stretch + noise filtering on mean NATO scale values.

PROCESSED SOFT-COPY INFORMATION EXTRACTION RESULTS

The processed soft-copy information extraction experiment was conducted using two Graeco-Latin squares as an experimental design. While this approach made maximum use of the limited PI resources available for the study, it also led to confounding of variable interactions with the effects of the image quality variables. Not surprisingly, the results of this study lacked in statistical power adequate to assess the effects of many of the experimental variables. Specifically, the combined blur/noise effect was statistically significant ($p < .01$), but the processes effect was not significant ($p = .10$). Further, the correlation between mean NATO scale values and percent correct EEIs was only .164 ($p = 0.28$). Thus, there is little value in the information extraction performance data from this experiment.

QUALITY METRIC EVALUATION

The various system metrics listed earlier were evaluated in terms of their ability to predict both NATO scale values and percent correct EEI responses. The image-dependent metrics were evaluated only in terms of their ability to predict NATO scaling values. Both system metrics and image-dependent metrics were calculated for both hard-copy and soft-copy experiments and correlated with EEI or NATO scale data. Because the system metrics were clearly nonlinearly related to both EEI and NATO scale data, a log-log transform was used to linearize the correlation. In the case of the image-dependent

metrics, the relationship was linear without a transformation.

The results of the correlation analyses between the metrics and experimental data are described below by metric category--system metrics and image dependent metrics.

System Image Quality Metrics

Product-moment correlations were obtained between the logarithm of the system metric and the logarithm of both EEI percent correct and mean NATO values. Subjective scaling and EEI performance data were averaged over all PIs and all 10 scenes to yield 25 (blur, noise combination) scores for both hard-copy and soft-copy display conditions. Details of these analyses are described by Beaton (1983) and are summarized in Tables 5 and 6.

As seen in Table 5, the correlations between metric values and percent correct EEIs were slightly higher for the soft-copy experiment than for the hard-copy experiment. While these differences were small in magnitude, it should be noted that all 16 metrics had higher correlations for the soft-copy than for the hard-copy data, a result which is highly significant ($p < .0001$). Thus, the system metric predictions are quite high in either case, predicting about 71 percent of the variance in EEI performance for the hard-copy case and 75 percent of the variance in soft-copy EEI scores. Because of the relatively small variation among the various metric correlations, it is difficult to single out a single metric as being superior based upon these results alone.

TABLE 5. Correlations between the $Logarithm_{10}$ of the Percent Correct EEIs and the $Logarithm_{10}$ of the System Metric Values

System Metric	Product-Moment Correlation Hard-Copy	Product-Moment Correlation Soft-Copy
EP	0.755	0.782
PEP	0.862	0.872
IR	0.755	0.785
PIR	0.922	0.944
SSF	0.756	0.787
PSSF	0.880	0.909
EW	-0.755	-0.785
PEW	-0.923	-0.944
MTFA	0.756	0.786
GSFP	0.773	0.801
ICS	0.922	0.944
VC	0.862	0.872
Q3	0.862	0.872
SN	0.930	0.952
PMQ	0.922	0.944
IC	0.825	0.854

On the average, the correlations between log NATO scale values and log system metric values are greater for the hard-copy imagery than for the soft-copy imagery, but this difference is again quite small. Some metrics are seen to be better predictors of hard-copy NATO values, while others better predict soft-copy NATO scores. In fact, exactly half the metrics predicted the hard-copy results better, while the other half predicted the soft-copy results with greater accuracy. There appear to be no strong trends among these data, although five of the metrics had correlations in excess of 0.90 for both hard- copy and soft-copy NATO scale prediction. Table 6 summarizes the results of these correlational analyses.

TABLE 6. Correlations between the Logarithm_{10} of the Mean NATO Scale Values and the Logarithm_{10} of the System Metric Values.

System Metric	Product-Moment Correlation Hard-Copy	Product-Moment Correlation Soft-Copy
EP	0.924	0.694
PEP	0.717	0.898
IR	0.924	0.705
PIR	0.903	0.951
SSF	0.925	0.712
PSSF	0.968	0.891
EW	-0.924	-0.705
PEW	-0.903	-0.951
MTFA	0.925	0.707
GSFP	0.938	0.735
ICS	0.903	0.951
VC	0.717	0.898
Q3	0.717	0.898
SN	0.895	0.948
PMQ	0.903	0.951
IC	0.964	0.827

Image-Dependent Quality Metrics

While the correlations in Tables 5 and 6 characterize image quality metrics based upon system capabilities, the metrics classified as image-dependent metrics take into account the image statistics of the individual scenes as displayed to the PI. For example, the image-dependent MTFA metric uses the modulation spectrum of an individual image cascaded with the system MTF to determine the area between the threshold curve and the displayed modulation spectrum.

Table 7 lists the resultant product-moment correlations for image-dependent metrics averaged across the 10 images for a total of 25 (blur X noise combination) data points. Correlations using all 250 images, without averaging across scenes, have also been calculated by Beaton (1983) and are consistently smaller in magnitude. Because some of the resulting correlations are not statistically significant, the associated probabilities of chance occurrence are also presented in Table 7.

As indicated in Table 7, the image-dependent metrics do not correlate as highly, in general, as do the system metrics. Many of the correlations are negative. While the soft-copy mean absolute correlation is higher than that for the hard-copy data, 5 of the 16 metrics predicted better for the hard-copy than for the soft-copy data.

TABLE 7. Correlations between the Mean NATO Scale Values and the Image-Dependent Metric Values

Image-Dependent Metric	Product-Moment Correlation	
	Hard-Copy	Soft-Copy
EP	0.569 ($p = .003$)	-0.136 ($p = .517$)
PEP	0.272 ($p = .188$)	0.508 ($p < .010$)
IR	-0.080 ($p = .704$)	-0.522 ($p = .008$)
PIR	0.554 ($p = .004$)	0.758 ($p < .001$)
SSF	-0.263 ($p = .204$)	-0.607 ($p = .001$)
PSSF	0.528 ($p < .007$)	0.688 ($p < .010$)
EW	-0.122 ($p = .562$)	0.271 ($p = .191$)
PEW	-0.407 ($p = .043$)	-0.734 ($p < .001$)
MTFA	0.840 ($p < .001$)	0.921 ($p < .001$)
GSFP	0.626 ($p < .001$)	0.794 ($p < .001$)
ICS	0.555 ($p < .004$)	0.759 ($p < .001$)
VC	0.272 ($p = .188$)	0.508 ($p = .009$)
Q3	0.269 ($p = .194$)	0.502 ($p = .011$)
PMR	0.557 ($p < .004$)	0.759 ($p < .001$)
PMQ	0.555 ($p = .004$)	0.759 ($p < .001$)
MSE	-0.733 ($p < .001$)	-0.368 ($p = .071$)

TABLE 7 (Continued)

PMSE	-0.528 ($p < .007$)	-0.199 ($p = .340$)
IF	0.868 ($p < .001$)	0.598 ($p < .002$)
CQ	-0.120 ($p = .568$)	0.278 ($p = .178$)
SC	0.548 ($p < .005$)	-0.186 ($p = .375$)
IC	0.572 ($p < .003$)	0.782 ($p < .001$)

To obtain a better indication of the relative prediction of the various image-dependent metrics, the magnitude of correlation was averaged algebraically for each metric across both hard- and soft-copy images and a relative ranking based on this correlation average was determined. (The metrics EW, PEW, MSE, and PMSE are expected, by their nature, to correlate negatively with performance, and therefore are treated as positive values for this purpose.) Those rankings are given in Table 8. It is of interest to note that the two best performing metrics, the MTFA and the GSFP, are the most evaluated metrics in the image quality literature, having been found to be robust in several experiments.

TABLE 8. Rank of Average Correlation by Metric Across Both Display Conditions

Metric	Rank	Metric	Rank
MTFA	1	SSF	12
IF	2	PEP	13
GSFP	3	VC	14
IC	4	Q3	15
PMR	5	PMSE	16
ICS	6	EP	17
PMQ	7	EW	18
PIR	8	SC	19
PSSF	9	CQ	20
PEW	10	IR	21
MSE	11		

V. DISCUSSION

As indicated above, the results contained in this report are merely an overview of the more important results of this entire research program and thus contain only the key points needed for such an overview. The interested reader or researcher is therefore advised to obtain copies of all the technical reports describing individual phases of the program to become familiar with specific details, more subtle results, and suggested applications. Nonetheless, several interesting and valuable issues have arisen from this research program and are noted in the results described above.

SCENARIO REALISM

While it is obviously impossible to conduct research of this nature which creates **precisely** the same problems for the PI as those which he/she experiences in a daily operational environment, we have been extremely pleased with the relative realism of the simulation and the acceptance in the intelligence community of our results. One original ground rule of the program was to use unclassified imagery, yet to make the content of that imagery and the various quality levels as close to operational levels as possible. From all discussions with many persons in the intelligence community, we believe we have succeeded in this area. The image content, while of domestic scenes, provided challenges similar to those encountered daily by operational PIs. The variety of image content covered representative orders of

battle (sea, air, land), while not favoring any particular scene content. The quality levels are considered to be quite representative of those of operational imaging systems. Lastly, the tasks required of the PIs (assigning NATO scale values and answering EEI questions) are precisely those performed on a routine basis; thus, there was no artificiality in the task for purposes of research simplification.

Because there was some concern about the validity of scoring of open-ended EEI questions, the early hard-copy study information extraction results were scored "blindly" by three separate individuals. Very high correlations were obtained between pairs of these individuals on individual images. Thus, it is believed that the development of the scenario, the EEIs, the use of the NATO scale, and the **a priori** specification of the EEI scoring criteria contribute a valuable addition to the experimental literature in the area of image interpretation. It is suggested that future researchers avail themselves of the database and the procedures followed in this program. Availability of the database is discussed in Appendix C.

HARD-COPY VS. SOFT-COPY INTERPRETATION

One of the major objectives of this program was to compare the efficacy of hard-copy imagery with that of soft-copy imagery. The first four experiments were designed to permit this direct comparison, both for subjective quality scaling and for information extraction. Comparing the results of these experiments, one finds that the results are somewhat mixed.

While the NATO scale values for soft-copy are typically higher than for hard-copy imagery presentation (Figures 14 - 16), the opposite conclusion is drawn for information extraction performance. That is, information extraction performance is consistently better for hard-copy than for soft-copy presentation, as indicated in Figures 17 - 19. While several explanations for these differences are reasonable, the following seems most likely. First, the PIs used in these experiments had no familiarity with soft-copy display of imagery, and therefore were both fascinated by it and enjoyed the manipulative capabilities of the soft-copy display. In addition, they were physically more comfortable looking at the soft-copy display than they were bent over a light table and looking through the fixed-position microscope. The comfort, paired with a "novelty" effect, probably resulted in increased subjective values of image quality for the soft-copy presentation.

On the other hand, the EEI scores are a measure of the actual performance of the PI using the imagery. There is no way that a novelty or preference effect can elevate these scores artificially, for the PI cannot obtain information from the image which is simply above the quality level of the image. For that reason, the EEI data are probably more objective in comparing the two presentation modes, leading to the conclusion that hard-copy interpretation is better than soft-copy interpretation, **under the conditions of these experiments**. This last point needs to be emphasized because the display used for soft-copy presentation was limited to 512 X 512 pixels and magnification was in discrete increments of 2X. Operational systems which have greater display information density (e.g., 1024 X 1024) or have continuous

zoom capability may produce different results. In fairness, it should be noted that the MTF of the display used in our soft-copy studies is as great or greater than that of many operational displays with better software capabilities.

While there appears to be a difference in actual EEI performance between the hard-copy and soft-copy presentations, favoring slightly the hard-copy mode, there is a very high correlation between NATO scale values and information extraction performance for both hard-copy and soft-copy experiments. These correlations are 0.898 for the hard-copy experiments and 0.965 for the soft-copy experiments, using the 15 blur/noise means collapsed across scenes in both cases. Thus, the behavior being measured by both subjective scaling and information extraction is highly correlated, permitting one to use scaling data (which are easier and more economical to acquire) for a variety of system evaluation and operational image screening purposes.

The more important consideration in selecting soft-copy presentation over hard-copy presentation is, of course, in the flexibility of image processing that is available from soft-copy. If the PI requires contrast modification, deblurring, or noise reduction in a hard-copy image, the computerized image (if available) must be manipulated and a new hard-copy print made. Contrast modification can of course be made in the darkroom without any computational capability, but even this requires substantial time (hours usually) compared with the seconds or minutes required for soft-copy processing and redisplay. Thus, a small penalty in soft-copy performance can easily be offset by more rapid processing and its attendant

improvement in image quality when using a soft-copy mode. The real question then lies in the efficacy of soft-copy image processing.

PROCESSED VS. NONPROCESSED SOFT-COPY INTERPRETATION

The inconsistency of the processed soft-copy EEI data was expected and is easily explained on the basis of the experimental design (Chao, 1983). Because there are high correlations between EEI performance and NATO scale values for both hard- and soft-copy nonprocessed conditions, it is reasonable to base conclusions regarding the efficacy of processed soft-copy presentation on the scaling data alone.

As illustrated in Figures 20 - 27, the various computer processes can produce a significant increase in subjective quality, often more than one NATO scale unit. Appropriately selected, the right process can result in improvements well in excess of the difference in scale value between the hard-copy and soft-copy conditions. Thus, the small loss in EEI performance found with the soft-copy presentation compared to the hard-copy presentation can be more than offset by the selection of the proper soft-copy process. In fact, the data suggest that the net benefit may be on the order of one-half to one full NATO scale value.

However, the problem of selecting the best computer process for soft-copy enhancement is not as simple as it might appear (Chao, 1983). The proper selection is certainly a function of **both** the blur and noise levels of the image, and may well depend somewhat upon the scene content. Of course, experience with particular processes and a

variety of scenes will enable the PI to select more efficiently the most useful process. Because the operational system will have a rapid response time (e.g., a few seconds) compared to this experimental system (2 - 120 s), sampling a few different processes may not be very inefficient. On the other hand, process selection must be done carefully to avoid performance degradation. There is no doubt that some of the claims made in the literature for some processes as being the panacea for all image quality ills are greatly exaggerated, if not utterly fallacious.

METRICS OF IMAGE QUALITY

Image quality metrics are desirable to have but often misleading. Various experiments have evaluated quality metrics *empirically* for television and hard-copy displays, but to our knowledge this is the first set of experiments designed to compare directly alternate metrics for both hard-copy and soft-copy imagery. The results of these comparisons are enlightening, but perhaps not totally conclusive. Without question, the MTFA measure performed well, as has been demonstrated frequently in the past (Borough, Fallis, Warnock, and Britt, 1967; Snyder, 1974, 1976; Task, 1976). Similar measures, such as the GSFP and IF performed nearly as well for image-dependent data. On the other hand, the MTFA did not perform as well as some other measures on a system basis, although nearly all of the measures were acceptable for system performance prediction.

Thus, it appears to be the case that overall, or average, system performance is easier to predict than is the PI's likely

performance with individual scenes. While the scene statistics provide the experimenter with considerable information from which to predict image quality, the metric does not weight the various areas of the scene in any cognitively relevant manner to permit the experimenter to obtain image statistics only from those relevant areas. Thus, the image statistics may contain a great deal of prediction "noise" which reduces the magnitude of the image-dependent predictions. Using a metric based only on overall system characteristics and ignoring specific scene characteristics is operationally useful but scientifically disappointing at this time.

One thing seems quite clear from the metric results. Those metrics which perform best take into account the perceptual limitations of the human observer, whereas those metrics that perform poorest are based largely on the image content or display system and do not weight any of the image information by the nonlinear sensitivity of the observer across the range of displayed spatial frequencies. A more detailed evaluation of the similarities and differences among the candidate metrics is offered by Beaton (1983), along with recommendations for metric selection.

VI. CONCLUSIONS

This research program has answered many of the questions it started out to answer. With direct regard to the experimental objectives of the program, the following results should be noted.

1. The experimental scenario that was developed and used in this research was operationally relevant, realistic, consistently valid, and capable of producing useful results both for basic research questions and for operational generalization. It is recommended for future researchers.

2. Soft-copy image displays are nearly as good as hard-copy displays for the same image quality. With an increase in displayed image content (more pixels per display) it is possible that this small difference will disappear. While the PI tends to believe that better image quality is seen on the soft-copy display, actual measurement confirms the opposite, although the difference is quite small.

3. Soft-copy processing can improve image quality, often as much as one NATO scale unit. Careful selection of the process is necessary, however, in that improper selection can degrade performance rather than improve it. The gains in interpretabilty through processing more than outweigh the losses in soft-copy display compared directly to hard-copy display.

4. Quality metrics can account for a great deal of the image quality variance and the variance in PI performance. Selection of the metric for overall system quality is easier than is selection for specific scene content.

REFERENCES

Beaton, R. J. Quality metrics of digitally derived imagery and their relation to interpreter performance: VII. Evaluation of quality metrics. Technical Report HFL-83-5, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Beaton, R. J., Chao, B. P., and Snyder, H. L. Quality metrics of digitally derived imagery and their relation to interpreter performance: IV. Subjective scaling of soft-copy digital imagery. Technical Report HFL-83-3, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Burke, J. J. and Strickland, R. L. Quality metrics of digitally derived imagery and their relation to interpreter performance: I. Preparation of a large-scale database. Technical Report 81-1, Optical Sciences Center, University of Arizona, Tucson, AZ, 1982.

Borough, H. C., Fallis, R. F., Warnock, R. H., and Britt, J. H. Quantitative determination of image quality. Boeing Company Technical Report D2-114058-1, 1967.

Chao, B. P. Quality metrics of digitally derived imagery and their relation to interpreter performance: VI. Evaluation of selected image processing algorithms. Technical Report HFL-83-4, Virginia

Polytechnic Institute and State University, Blacksburg, VA, 1983.

Chao, B. P., Beaton, R. J., and Snyder, H. L. Quality metrics of digitally derived imagery and their relation to interpreter performance: V. Effects of blur and noise on soft-copy interpretability. Technical Report HFL-83-2, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Snyder, H. L. Image quality and face recognition on a television display. **Human Factors**, 1974, **16**, 300-307.

Snyder, H. L. Visual search and image quality. U.S. Air Force Technical Report AMRL-TR-76-89, 1976.

Snyder, H. L., Shedivy, D. I., and Maddox, M. E. Quality metrics of digitally derived imagery and their relation to interpreter performance: III. Subjective scaling of hard-copy digital imagery. Technical Report HFL-81-3, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1982.

Snyder, H. L., Turpin, J. A., and Maddox, M. E. Quality metrics of digitally derived imagery and their relation to interpreter performance: II. Effects of blur and noise on hard-copy interpretability. Technical Report HFL-81-2, Virginia Polytechnic Institute and State University, Blacksburg, VA. 1982.

Task, H. L. An evaluation and comparison of several measures of image

quality for television displays. U.S. Air Force Technical Report AMRL-TR-79-7, 1979.

APPENDIX A: PRESENTATIONS AND PUBLICATIONS

Beaton, R. J. Quality metrics of digitally derived imagery and their relation to interpreter performance: VII. Evaluation of quality metrics. Technical Report HFL-83-5, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Beaton, R. J. Quantitative models of image quality. **Proceedings of the Human Factors Society**, 1983, in press.

Beaton, R. J., Chao, B. P., and Snyder, H. L. Image interpretation for soft-copy displays of digitally derived images. **Proceedings of the Human Factors Society**, 1982, 461-464.

Beaton. R. J., Chao, B. P., and Snyder, H. L. Quality metrics of digitally derived imagery and their relation to interpreter performance: IV. Subjective scaling of soft-copy digital imagery. Technical Report HFL-83-3, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Beaton, R. J., Monty, R. W., and Snyder, H. L. An evaluation of system quality metrics for hard-copy and soft-copy displays of digital imagery. **Proceedings of the SPIE**, 1983, in press.

Burke, J. J. and Snyder, H. L. Quality metrics for digitally derived imagery and their relation to operator performance: I. Preparation of a large-scale database. Invited paper in P. N. Slater and R. F.

Wagner (Ed.), **Image Analysis Techniques and Applications**, Society of Photographic Scientists and Engineers, Washington, D. C., January 1981.

Burke, J. J. and Snyder, H. L. Quality metrics of digitally derived imagery and their relation to interpreter performance. **Proceedings of the SPIE, 310**, August 1981.

Burke, J. J. and Strickland, R. N. Quality metrics of digitally derived imagery and their relation to interpreter performance: I. Preparation of a large-scale database. Technical Report HFL-81-1, Optical Sciences Center, University of Arizona, Tucson, AZ, 1982.

Chao, B. P. Performance evaluations of digital image quality processed by enhancement techniques. **Proceedings of the Human Factors Society**, 1983, in press.

Chao, B. P. Quality metrics of digitally derived imagery and their relation to interpreter performance: VI. Evaluation of selected image processing algorithms. Technical Report HFL-83-4, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Chao, B. P., Beaton. R. J., and Snyder, H. L. Human performance evaluations of digital image quality. **Proceedings, Advances in Display Technology** III, SPIE--The International Society for Optical Engineering, 1983, **386**, 20-24.

Chao, B. P., Beaton, R. J., and Snyder, H. L. Evaluation of CRT-displayed digital imagery using subjective scaling. **Proceedings of the Human Factors Society**, 1982, 329-333.

Chao, B. P., Beaton. R. J., and Snyder, H. L. Quality metrics of digitally derived imagery and their relation to interpreter performance. V. Effects of blur and noise on soft-copy interpretability. Technical Report HFL-83-2, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Shedivy, D. I. and Snyder, H. L. Subjective scaling of digital imagery. **Proceedings of the Human Factors Society**, 1981, 30-32.

Snyder, H. L. and Burke, J. J. Quality metrics for digitally derived imagery and their relation to operator performance: II. Interpretability and judged quality of hard-copy imagery. Invited paper in P. N. Slater and R. F. Wagner (Ed.), **Image Analysis Techniques and Applications**, Society of Photographic Scientists and Engineers, Washington, D. C., January 1981.

Snyder, H. L., Shedivy, D. I., and Maddox, M. E. Quality metrics of digitally derived imagery and their relation to interpreter performance: III. Subjective scaling of hard-copy digital imagery. Technical Report HFL-81-3, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

Snyder, H. L., Shedivy, D. I., Turpin, J. A., Maddox, M. E., Burke, J.

J., and Strickland, R. N. Digital image quality and interpretability: database and hardcopy studies. **Optical Engineering**, 1982, **21**(1), 14-23.

Snyder, H. L., Turpin, J. A., and Maddox, M. E. Quality metrics of digitally derived imagery and their relation to interpreter performance: II. Effects of blur and noise on hard-copy interpretability. Technical Report HFL-81-2, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1982.

Strickland, R. N. and Burke, J. J. Digital image quality and interpretability: the efficiency of enhancement and restoration algorithms. **Proceedings of the IEE Conference on Electronic Image Processing**, York, Great Britain, July 1982.

Turpin, J. A. Effects of noise and blur on digital imagery interpretation. **Proceedings of the Human Factors Society**, 1981, 286-289.

APPENDIX B: STAFF AND STUDENT PARTICIPATION IN THIS RESEARCH

The following research staff and faculty were fully or partially supported by this research effort during the periods of time indicated.

Harry L. Snyder, Professor, VPI&SU (1978-1983)
James J. Burke, Professor, UA (1978-1983)
Charles D. Bernard, Research Associate, VPI&SU (1978-1980)
Michael E. Maddox, Research Associate, VPI&SU (1978-1980)
L. Hardy Mason, Research Associate, VPI&SU (1980-1982)
Robin N. Strickland, Research Associate, UA (1978-1982)

The following graduate students worked on this project, receiving the degrees indicated as part of their research activity.

Robert J. Beaton, Ph.D. (1983), VPI&SU
Thomas J. Bruegge, M. S. (1982), UA
Kenneth E. Castle, M. S. (1981), Ph.D. (exp. 1984), UA
Betty P. Chao, Ph.D. (1983), VPI&SU
Robert W. Monty, M. S. (exp. 1984), VPI&SU
David I. Shedivy, M. S. (1981), VPI&SU
James A. Turpin, M. S. (1981), VPI&SU

APPENDIX C: AVAILABILITY OF THE DATABASE

The database used in this research effort was developed at considerable expense for both hard-copy and soft-copy experimentation. The soft-copy database exists in standard IBM 9-track, 800 bpi magnetic tape format. It is available to qualified users for the cost of copying the tapes. In the research conducted to date none of the images have been published to avoid contamination of the results of possible future experiments from knowledge or viewing of the images by potential subjects. Thus, the intent of the researchers is to avoid contamination for future experiments by careful screening of users and recipients of the database.

APPENDIX D: THE NATO SCALE

Rating Category 0

Useless for interpretation due to cloud cover, poor resolution, etc.

Rating Category 1

Detect the presence of larger aircraft at an airfield.

Detect surface ships.

Detect ports and harbors (including piers and harbors).

Detect railroad yards and shops.

Detect coasts and landing beaches.

Detect surfaced submarines.

Detect armored artillery ground force training areas.

Recognize urban areas.

Recognize terrain.

Rating Category 2

Detect bridges.

Detect ground forces installations (including training areas, administration/barracks buildings, vehicle storage buildings, and vehicle parking areas).

Detect airfield facilities (count accurately all larger aircraft, by type, straight-wing and swept/delta wing).

Recognize ports and harbors (including large ships and drydocks).

Rating Category 3

Detect communications equipment (radio/radar).

Detect supply dumps (POL/ordnance).

Detect and count accurately all straight-wing aircraft, all swept-wing aircraft, and all delta-wing aircraft.

Detect command and control headquarters.

Detect surface-to-surface and surface-to-air missile sites (including vehicles and other pieces of equipment).

Detect land minefields.

Recognize bridges.

Recognize surface ships (distinguish between a cruiser and a destroyer by relative size and hull shape).

Recognize coast and landing beaches.

Recognize railroad yards and shops.

Recognize surfaced submarines.

Identify airfield facilities.

Identify urban areas.

Identify terrain.

Rating Category 4

Detect rockets and artillery.

Recognize troop units.

Recognize aircraft (such as FAGOT/MIDGET when singly deployed).

Recognize missile sites (SSM/SAM). Distinguish between missile types by the presence and relative position of wings and

control fins.

Recognize nuclear weapons components.

Recognize land minefields.

Identify ports and harbors.

Identify railroad yards and shops.

Identify trucks at ground force installations as cargo, flatbed, or van.

Identify a KRESTA by the helicopter platform flush with the fantail, a KRESTA II by the raised helicopter platform (one deck level above fantail and flush with the main deck).

Rating Category 5

Detect the presence of call letters or numbers and alphabetical country designators on the wings of large commercial or cargo aircraft (where alphanumerics are three feet high or greater).

Recognize command and control headquarters.

Identify a singly deployed tank at a ground forces installation as light or medium/heavy.

Perform Technical Analysis (PTA) on airfield facilities.

PTA on urban areas and terrain.

Rating Category 6

Recognize radio/radar equipment.

Recognize supply dumps (POL/ordnance).

Recognize rockets and artillery.

Identify bridges.

Identify troop units.

Identify coast and landing beaches.

Identify a FAGOT or MIDGET by canopy configuration when singly deployed.

Identify the ground force equipment T-54/55 tank, BTR-50 armored personnel carrier, or 57 mm AA gun.

Identify, by type, RBU installations (e.g., 2500 series), torpedo tubes (e.g., 21 inch/53.34 cm), and surface-to-air missile launchers on a KANIN DDG, KRIVAC DDGSP, or KRESTA II.

Identify a ROMEO-class submarine by the presence of the cowling for the snorkel induction and the snorkel exhaust.

Identify a WHISKEY-class submarine by the absence of the cowling and exhaust.

Rating Category 7

Identify radar equipment.

Identify major electronics by type on a KILDEN DDGS or KASHIN DLG.

Identify command and control headquarters.

Identify nuclear weapons components.

Identify land minefields.

Identify the general configuration of an SSBN/SSGN submarine sail, to include relative placement of bridge periscope(s) and main electronics/navigation equipment.

PTA on ports, harbors, and roads.

PTA on railroad yards and shops.

Rating Category 8

Identify supply dumps (POL/ordnance).

Identify rockets and artillery.

Identify aircraft.

Identify missile sites (SSM/SAM).

Identify surface ships.

Identify vehicles.

Identify surfaced submarines (including components such as ECHO II SSGN sail missile launcher elevator guide and major electronics/navigation equipment by type).

Identify, on a KRESTA II, the configuration of the major components of larger electronics equipment and smaller electronics by type.

Identify limbs (arms, legs) on an individual.

PTA on bridges.

PTA on troop units.

PTA on coast and landing beaches.

Rating Category 9

Identify in detail the configuration of a D-30 howitzer muzzle brake.

Identify in detail on a KILDEN DDGS the configuration of torpedo tubes and AA gun mountings (including gun details).

Identify in detail the configuration of an ECHO II SSGN sail including detailed configuration of electronics communications equipment and navigation equipment.

PTA on radio/radar equipment.

PTA on supply dumps (POL/ordnance).

PTA on missile sites.

PTA on nuclear weapons components.